

10/070386

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 09:32:10 ; Search time 2707.4 Seconds
(without alignments)

15199.092 Million cell updates/sec

Title: US-10-070-386-1

Perfect score: 1378

Sequence: 1 gtcagcgggtgggtgatat.....accatcaatcaactaaca 1378

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estm.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hic.*

9: gb_est1.*

10: gb_est2.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
C 1	43.8	3.2	885	13	BX425603
C 2	42.4	3.1	1201	13	BX376097
C 3	41.6	3.0	557	28	BH285801
C 4	39	2.8	1031	29	CNS05QW

RESULT 1

BX425603/c

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

ALIGNMENTS

885 bp mRNA linear EST 15-MAY-2003

NEUROBLASTOMA Homo sapiens cDNA clone

3-PRIME, mRNA sequence.

EST

GI:30770486

Human sapiens (human)

Homo sapiens

Sukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Li, W.B., Gruber, C., Jesse, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/invitrogen/Corporation 1600

Paradise Avenue Genoscope sequence ID : CLOBB022ZA07P1.

Location/Qualifiers

1. .885

/organism="Homo sapiens"

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CL08B0222A07"
/tissue_lib="Homo sapiens NEUROBLASTOMA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 3.2%; Score 43.8; DB 13; Length 885;
Best Local Similarity 3.3%; Pred. No. 0.24;
Matches 15; Conservative 146; Mismatches 293; Indels 0; Gaps 0;
QY 92 AAGACTGAAGAGGACCAAGATATTCACATGCGATACGGTGTGATCATGCTGTTT 151
DB MAMMM 701
QY 152 CAAGAGAGACGCTTCATCTACCTGCTGCTCCCTCTCTGTTGTACAGATCAAGTATCG 211
DB MAMMM 641
QY 212 GATGACACCCCGCAAGATTCCTGGAGTTCAAGAGGGTGTGCTCTACGCGCATTT 271
DB MAMMM 581
QY 272 AGTATAGATGTCATGAGGTTGACGTAAGCTGAAGCTGATTAAGAGATGAGACAC 331
DB MAMMM 521
QY 332 GAAATAACAGGTTGATGCTTCCCTGCTTACTAAAGTGAATCAAGAGACACACA 391
DB MAMMM 461
QY 392 GCCGAAGAACCGATGCTGCTGAGGGGTTCCCTTACAGTCTACATGTAACGTCAT 451
DB MAMMM 401
QY 452 GATAGAAACATCAATGCGCAATAGTGTAGTATACCTGACGCTACATGCTTCTCCG 511
DB MAMMM 341
QY 512 GATCTGCTAAATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 545
DB MAMMM 307

RESULT 2

LOCUS BX376097/c
DEFINITION BX376097 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens cDNA clone CSODC022TM12 5-PRIME, mRNA sequence.

ACCESSION BX376097
VERSION BX376097.1 GI:30434756

KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1201)

Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seq@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 2866.f

Contact: Feng Liang Email: fliang@lifetech.com URL:

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Paraday Avenue Genoscope sequence ID : CSODC022BG06QP1.

FEATURES

source

1. 1201

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSODC022TM12"

/tissue_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 3.1%; Score 42.4; DB 13; Length 1201;

Best Local Similarity 12.1%; Pred. No. 0.74; Indels 1; Gaps 1;

Matches 89; Conservative 260; Mismatches 385; Indels 1; Gaps 1;

QY 297 GTAAGCTGAAGAGCTGATTACGAGACATGACACCAACGAAATACACCGTTGTATGCGTTC 356

DB GKRGGKGGDAADAAAAAAGKAAADAKGKAGKAAAGAAAGAAAGAAAGAAAGAAAG 1128

QY 357 CCGTGTCTTAAAGTATATCCAGAGCAACACAGCCGAAAGAACCGGATGCTGTCTGA 416

DB KARGAAAGAGKAGAGAGGAGGARRKAKAGAGTARWGTAAAGKARAKARAKARAKADAA 1068

QY 417 GGGGTTCTTTAGAGTCTACATGCTAACGCTGATAGAAACATCAAAATGCCCAATCA 476

DB DGGKAAKAGKAG 1008

QY 477 AGTTAGTATACGCTACATCGCTTCTTCGCGATCTTGCCTAAATATATATGTCCT 536

DB MAMMM 948

QY 537 GTCCGAATCTCGTACTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 596

DB MAMMM 888

QY 597 GCCCGCTTTGAGACCTACATGATGACATCTTAAGAGAGGATCTGACACATTTCTA 656

DB GGMATVGMGGVGMGGTGMGGTGMGGTGMGGTGMGGTGMGGTGMGGTGMGGTGMGGTGM 828

QY 657 AGGCATCCATATAGGCATTTGGGGCTTAAGTCGCAATTAAGAGAGATTAAGGGGGTGTGAA 716

DB NGKTMHGKNNMM 768

QY 717 AGTGGTGTGCAAAAGAGGATCGATTCGCTATACCGCCGCTAAGAGGCTGGGCTAGCAG 776

DB KMGKTMGTMMGTGMGGTGMGGTGMGGTGMGGTGMGGTGMGGTGMGGTGMGGTGMGGTGM 708

QY 777 CTGCTCGACCTGGAATACGTCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 836

DB MAMMM 648

QY 837 CAAATGCTGATTCGGTTTAAATGGGATGTAGTGTAGGTGCGGAAACACAGTTTATGATCT 896

DB MAMMM 589

QY 897 AGTTAAAGGGAAGCTGAAAGCTGAACTGTGCAAGAAATAGCCTGTTGGATATCAAGCTTG 956

DB MAMMM 529

QY 957 ATAACCAATTCAGTGTGCTCAAGGGTGTCTCTGATATGCTGGAGCTTCCTGCTGCGATG 1016

DB MAMMM 469

QY 1017 GGGTAACATATTCAT 1031

DB MAMMM 454

RESULT 3

BH285801
LOCUS 557 bp DNA linear GSS 30-NOV-2001
DEFINITION CH230-118D10.TVB CHORI-230 segment 1 Rattus norvegicus genomic
clone CH230-118D10, genomic survey sequence.
ACCESSION BH285801
VERSION BH285801.1 GI:17198207
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 557)
Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,
Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D.,
Rat, F.P., de Jong, P. and Fraser, C.M.
Rat BAC End sequences from Library CHORI-230 EcoRI segment
Unpublished (1999)
Other GSSs: CH230-118D10.TJB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(<http://www.chori.org/bacpac/orering-information.htm>). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 118 row: D column: 10
Seq primer: T7
Class: BAC ends.

```

FEATURES
    source
    class, acc ends,
    Location/Qualifiers
        1..557
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /strain="BN/SsHsd/MCW"
            /db_xref="taxon:10116"
            /clone="CH230-118D10"
            /sex="Female"
            /cell_type="Brain"
            /clone_lib="CHORI-230 Segment 1"
            /notes="Vector: pTARBAC2.1; Site 1: EcoRI;
CHORI-230 Rat (BN/SsHsd/MCW) BAC library produced by
Pieter de Jong"

```

ORIGIN		Fischer de Young	
Query Match	3.0%;	Score 41.6;	DB 28; Length 557;
Best Local Similarity	49.1%;	Pred. No. 0.85;	
Matches 110;	Conservative 0;	Mismatches 114;	Indels 0; Gaps 0;
Qy	1155	AGAGTGGGCAATTGAAGTAGCTACCGCATCGAACCCGAGCGTTCAAGACATGGCGCT	1214
Db	89	AGATGCGCTGGAAATTAAGAGAAATGACAAGAAGAGGGGGAGACATATGAAGAGTCATCAGG	148
Qy	1215	ACGTAGATACATAGAGTGTCATAGAAACATAAAGAGAGGCTTGAAGAACCATTCAAATCCTAA	1274
Db	149	ACATGTATGCACATATTACTAGCTAGCTTCAAGAAGCATCATTTCTATTAAATATTATAA	208
Qy	1275	GGGTCTCTCTCTTCTGTGCATCATCAATCAAGATCATACACTCCTCAACACAGAACTTTCT	1334
Db	209	CGTTATCATGTTTTTCTAAAAAATAAATAAGTTAAAGTACAGAGNACTCTATGA	268
Qy	1335	ATCTCCCTATAGCAATTCCTCAAAACCCATCAATCAACCTTAACA	1378
Db	269	GACTGAACTAGAGAAATCCCCAAACCTCTTAGAGATACATAGA	312

RESULT 4
CNS05QQW/c

LOCUS	CNS050QW	1031 bp	DNA	linear	GSS 01-SEP-2000
DEFINITION	Tetraodon nigroviridis genome survey sequence SP6 end of clone 025803 of library B from Tetraodon nigroviridis, genomic survey sequence.				
ACCESSION	AL349601.1 GI:8243371				
VERSION	GSS; genome survey sequence.				
KEYWORDS	Tetraodon nigroviridis				
SOURCE	Tetraodon nigroviridis				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontoidea; Tetraodontidae; Tetraodon.				
REFERENCE	1 Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Winkler,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J. Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence Nat. Genet. 25 (2), 235-238 (2000)				
JOURNAL	20296633				
MEDLINE	10835645				
PUBMED					
REFERENCE	2 Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis Genome Res. 10 (7), 939-949 (2000)				
JOURNAL	20359837				
MEDLINE	10899143				
PUBMED					
REFERENCE	3 (bases 1 to 1031) Direct Submission Genoscope.				
JOURNAL	Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : EP 191 91006 EVRY Cedex - FRANCE [E-mail : seqref@genoscope.cns.fr] - Web : www.genoscope.cns.fr				
MEDLINE	Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : EP 191 91006 EVRY Cedex - FRANCE [E-mail : seqref@genoscope.cns.fr] - Web : www.genoscope.cns.fr				
PUBMED					
REFERENCE	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .				
JOURNAL					
MEDLINE					
PUBMED					
FEATURES	Location/Qualifiers				
source	1..1031 /organism="Tetraodon nigroviridis" /mol_type="genomic DNA" /db_xref="taxon:99983" /clone="025803" /clone_lib="B" /note="Genoscope sequence ID : COAB025CA02B1-end : SP6"				
ORIGIN					
Query Match	2.8%; Score 39; DB 29; Length 1031;				
Best Local Similarity	43.1%; Pred. No. 7.2;				
Matches	81; Conservative 0; Mismatches 107; Indels 0; Gaps 0;				
QY	792 AATAAGCTCAGTCTGCTAGGTATGTCACCTTAATGTCACAGATGCAATGCTGATTGGG 851				
Db	523 AAAAAAGTTTGTCTGCTAGCAGCTNACTANAAATTNACNAGANAANAAGNAATAAGT 464				
QY	852 TTAAATGGGCACTGATGTAGTGGTCCGAAAACACGTTTAGATCTAGTTAAAGGGAGCT 911				
Db	463 NTNAAATTGCGANNNNNNNTATTNATGTCATNAGNANATNTGTNGCTNATATNAGGTAGTAA 404				
QY	912 GAAAGCTGAACCTGTCAGAAATAAGCGCTGTGGAATACAAACGTTTGATATCCCAATTCAGT 971				
Db	403 NAATNCCTACCATNTNAAAACACNGNANTNATGNANACTNAGTTCCAAATCCNACTCTTT 344				
QY	972 CGTCAAGG 979				
Db	343 CTTNAGG 336				

RESULT 5

```

BX381961      1201 bp      mRNA      linear      EST 08-MAY-2003
LOCUS      BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION      clone CSODI072YF05 3-PRIME, mRNA sequence.
ACCESSION      BX381961
VERSION      BX381961.1 GI:30453007
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      1 (bases 1 to 1201)
JOURNAL      Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
COMMENT      Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CSODI072CC03NP1.
Location/Qualifiers
FEATURES             source
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI072YF05"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match      2.8%; Score 39; DB 13; Length 1201;
Best Local Similarity 3.5%; Pred.No.7.8;
Matches 18; Conservative 159; Mismatches 343; Indels 0; Gaps 0;

QY      687 CGGCAATGAAGAGATAGGGGGGTGTGAAGTGGTGTGTCACAAAGGAGGTGCGATGGCT 746
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      385 CKANNKKKMKACVANNNNKCKMMNNNNKKKKKKMMKKKKKKKKKKKKKKKKKKKK 444
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      747 ATACGAGCGGTAGCAGGTGGGTAGCAGGTGTCTGAGCTGTCTGAGCTGTGAATACGTCATGCG 806
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

DB      445 NNNCNVANNMMNNKNNKNNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 504
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      807 TTAGGTATGTCACCTAATGTGACAGATGCAATGTCTGATGGTGTAAATGGGCATGT 866
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

DB      505 CNNKKKCMNNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 564
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      867 AGTGTAGGTGCGAAACACGTTTAGATCTTAGTTAAAGGGGAGCTGAAAGCTGAACCTGT 926
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

DB      565 NNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 624
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      927 CAGAAATAAGCTGTGGAATAACAGCTGTGATACCACTTCACTCGTCAAGGCTGTCT 986
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

DB      625 KMMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 684
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      987 GATATGTGGAGCTTCCCTGTGCGATGTGTGGGGTAACTATTTCATAGTGGGCGAGATGC 1046
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

DB      685 KNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 744
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      1047 AACTCTATTTCATTTGAATTAACCTATTCTGGGTAGGATGCTCAATGCTTCTTCGC 1106
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

DB      745 NCNKKNNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 804
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      1107 TGTCACTTACACATATGCGGGGTCAACACAGCTATACAGCTTCATAGAGAGTGCAGCAT 1166
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

DB      805 KKKKKMBKGMVCKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 864

```

```

QY      1167 TGAAGTAGTACGCGATCGACCCCGAGAGCGGTTCAGAC 1206
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      865 NKAMMKKDNMMNNMMKMMCMKMMKMMKMMKMMKMMKMMKMMKMMKMMKMMKMM 904
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
LOCUS      CC264185      1205 bp      DNA      linear      GSS 13-MAY-2003
DEFINITION      CH261-57E7_Sp6.1 CH261 Gallus gallus genomic clone CH261-57E7,
genomic survey sequence.
ACCESSION      CC264185
VERSION      CC264185
KEYWORDS      GSS.
SOURCE      Gallus gallus (chicken)
ORGANISM      Gallus gallus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
TITLE      1 (bases 1 to 1205)
JOURNAL      Kremitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
COMMENT      Warren, W., Graves, T., Mardis, E. and Wilson, R.
Unpublished (2003)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 182000 Std Error: 0.00
Seq primer: Sp6 ATTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 50
High quality sequence stop: 729.
Location/Qualifiers
FEATURES             source
1..1205
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strains="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-57E7"
/sex="female"
/cell_line="UCD001, inbred 256"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - For library and clone
ordering information: http://www.chori.org/bacpac"
ORIGIN
Query Match      2.8%; Score 38.8; DB 28; Length 1205;
Best Local Similarity 48.2%; Pred.No.9;
Matches 109; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY      823 AATGTCAGCAGATGCAATGCTGATGGGTAAATGGCGATGTAGTGTGCGGAAA 882
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      845 AATCTCCACTGACAGATTGATTATGGGATGAAGGGAAGGCTTGTGGGGCGAGGAA 904
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      883 ACACGTTTAGATCTAGTTAAAGGGAGCTGAAAGCTGAAACCTGTCTAGAAATAAGCTGTT 942
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      905 AAAAGTTAAACCCCACTTATAAATGAACCCCTCAGTAAACCTTAAAGATCGAACCAAGA 964
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      943 GGAATACACAGTTGATACCCATTCAGTCGTCAGGGGTCTCTGTATATGCTGAGCTTTC 1002
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      965 AGAATAAAGGGGGTCTATCAACCTTGTGTATGAAGGGGTATCCCATTAATGTTTCCITTA 1024
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      1003 CTGTGCGCATTTGGGGTAACTATTTCATAGTGGGGCAGAAATGCAA 1048
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1025 ACCTTCATTCGGGCAACAAAATTCATGCGAGGGGAGAGAAATAA 1070
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
LOCUS      BZ246183/c      673 bp      DNA      linear      GSS 12-OCT-2002
DEFINITION      CH230-374A18.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-374A18, genomic survey sequence.

```

```

ACCESSION   B2246183
VERSION     B2246183.1  GI:23906447
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

REFERENCE   1 (bases 1 to 673)
AUTHORS    Zhao S., Shetty J., Shatsman S., Teagay G., Geer K.,
            Shwartsbeyn A., Gebregeorgis E., Overton L., Russell D., Chen D.,
            Riggs P., de Jong P. and Fraser C.M.
TITLE      Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL    Unpublished (1999)
COMMENT    Other GSSs: CH230-374A18.TU
            Contact: Shaving Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.html). For BAC library
            availability, please contact Pieter de Jong (pjejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering information.htm). BAC end
            page: http://www.tigr.org/tcdb/bac_end/rat/bac_end_intro.html
            Plate: 374 row: A column: 18
            Seq primer: T7
            Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..673
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsHsd/MCw"
                     /db_xref="taxon:10115"
                     /clone="CH230-374A18"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /notes="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
                     CHORI-230 Rat (BN/SsHsd/MCw) BAC library produced by
                     Pieter de Jong"

ORIGIN
Query Match       2.8%; Score 38.4; DB 28; Length 673;
Best Local Similarity 51.1%; Pred. No. 8.6;
Matches 90; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Qy 1100 TCTCGCTGTCACTTACACATCATCGGGGTCAACACGATATACAGCTTCATAGAGT 1159
      |||||
Db 506 TCTGTTTAAATTAATAGAAATGCTATTATATATATATATATATATATATATATATACACAGAGA 547

Qy 1160 GCGGATTGAAGTAGCTACCCATCGAACCCGGAAGCGGTTCAAGACATGGCGTAGCTA 1219
      |||||
Db 546 GAGACACAGACAGACAGACACACAGACACAGACACAGACAGACACAGACACAGACAGACAG 487

Qy 1220 GATACATAGAGTCATAGAAATTAAGAGGAGCTTGAAGACCATTCAAATCCTAAG 1275
      |||||
Db 486 CACAGATACACACAGACACACAGAGGGAGATTCTCGAAGCTTCTCAAGAGAAAG 431

RESULT 8
ACQ18895/c
LOCUS       ACQ18895
DEFINITION  CIT-HSP-2302116.TF CIT-HSP Homo sapiens genomic clone 2302116,
            genomic survey sequence.
ACCESSION   ACQ18895
VERSION     ACQ18895.1  GI:3197631
KEYWORDS    Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM    Homo sapiens

Query Match       2.8%; Score 38.2; DB 28; Length 479;
Best Local Similarity 50.3%; Pred. No. 8.2;
Matches 94; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 31 GTGTCCAAAACCTGTGTAGTGACTACGAATGAGGAAAGAAACGGTGTGTGTGGCAGCT 90
Db 460 GTGTACAAATGCTAAAGAAATTTTGAAGAAAGAAAGAAATAGTTGAATGTGGCAGGG 401

Qy 91 GAGACTGAGAGAGGAGCCAAAGATATTCACATGCGATACGGTTCATCATGCTTGT 150
      |||||
Db 400 AGTGAAGAAAGTAGAGGGTATAGATAATATGAATGTTATATATCGGTGTGGC 341

Qy 151 TCAAGAGAGACACCGTTCATCTACCTGGTGTTCCTCTTCCTGTTGTACAGATCAAGTATC 210
      |||||
Db 340 TAGGCAATGGACATGATGATGACATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 281

Qy 211 GGATGAC 217
      |||||
Db 280 GGAATAC 274

RESULT 9
ACQ1895/c
LOCUS       ACQ1895
DEFINITION  CIT-HSP-2302116.TF CIT-HSP Homo sapiens genomic clone 2302116,
            genomic survey sequence.
ACCESSION   ACQ1895
VERSION     ACQ1895.1  GI:8038159
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
ORGANISM    Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodontidae; Tetraodon.
            Roest Crolius H., Jaillon O., Dasilva C., Bouneau L., Fisher C.,

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 479)
Adams M.D., Rounsley S.D., Zhao S., Field C.E., Bass S., Linher K.,
Golden K., Berry K., Granger D., Suh E., Wible C., Shizuya H.,
Simon M. and Venter J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
Other GSSs: CIT-HSP-2302116.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tcdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..479
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2302116"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /notes="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
                     HindIII"

ORIGIN
Query Match       2.8%; Score 38.2; DB 28; Length 479;
Best Local Similarity 50.3%; Pred. No. 8.2;
Matches 94; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 31 GTGTCCAAAACCTGTGTAGTGACTACGAATGAGGAAAGAAACGGTGTGTGTGGCAGCT 90
Db 460 GTGTACAAATGCTAAAGAAATTTTGAAGAAAGAAAGAAATAGTTGAATGTGGCAGGG 401

Qy 91 GAGACTGAGAGAGGAGCCAAAGATATTCACATGCGATACGGTTCATCATGCTTGT 150
      |||||
Db 400 AGTGAAGAAAGTAGAGGGTATAGATAATATGAATGTTATATATCGGTGTGGC 341

Qy 151 TCAAGAGAGACACCGTTCATCTACCTGGTGTTCCTCTTCCTGTTGTACAGATCAAGTATC 210
      |||||
Db 340 TAGGCAATGGACATGATGATGACATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 281

Qy 211 GGATGAC 217
      |||||
Db 280 GGAATAC 274

RESULT 9
ACQ1895/c
LOCUS       ACQ1895
DEFINITION  CIT-HSP-2302116.TF CIT-HSP Homo sapiens genomic clone 2302116,
            genomic survey sequence.
ACCESSION   ACQ1895
VERSION     ACQ1895.1  GI:8038159
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
ORGANISM    Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodontidae; Tetraodon.
            Roest Crolius H., Jaillon O., Dasilva C., Bouneau L., Fisher C.,

```

Nat. Biotechnol. 19 (5), 440-445 (2001)
21327151
11329013
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave., Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@atersys.com
High quality sequence ston: 451.

Location/Qualifiers
1. 482

```
/mol_type="RNA"  
/db_xref="taxon:9606"  
/cell_line="HT1080"
```

```
/note="See 'Creation of  
Libraries using Random A
```

method was used, these expressed in HT1080 under

ity 55.5%; Pred. No. 9
servative 0; Mismatch

GAACTGTCTCAGAAATAAGCCTGTTCTC

GTGTTCCCTCCCACTCNAACAGCC

15 JAN 1964

1201 bp
Homo sapiens FETAL BRAIN

..1 GI:30654824

a; Metazoa; Chordata; Cr
Eutheria; Primates; Cat

length cDNA libraries and normalized (2001)

1006 EVRY cedex - France
greg@genoscope.cns.fr, w

www.genoscope.cns.fr/

Avantec Genoscope sequences

/ORGANISM="HOMO SAPIENS"

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODF020YA02"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NciI-oligo (GT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN

Query Match      2.8%; Score 38; DB 13; Length 1201;
Best Local Similarity 17.4%; Pred. No. 16;
Matches 65; Conservative 137; Mismatches 172; Indels 0; Gaps 0;

Qy 665 ATATAGGCGATTGGCGGCTTAAGTCGGCATTTGAAGAGATAAGGGGGTGTGAAAGTGCTGT 724
Db 741 RTKGGKWTGKGGGGGGGGGATTTAGGGGKTKARGAWAGDKGGTTTAGDITKK 800

Qy 725 GTCAAAAGGAGGTGATTGGCTATACACAGCGCTAAGCAGGTGGGCTAGCAGCTGTCTGC 784
Db 801 AGAKGWRGTGKTAKAGKGTAWGGKGGKATWGGGGGGGGKAKDGGGGGTTAGA 860

Qy 785 AGCTGTGAATAACGTCCTCTAGGTATGTCACCTTAATGTCAGCAGATGCAAAATGCT 844
Db 861 AGGTTGGGGTGDAATGGKGGKGGKRAAAGKTTTKTAADKKKTKTKTAAXKKEK 920

Qy 845 GATTGGGTTAAATGGGCGATAGTGTAGGTGGCGGAAACACAGTTTATAGTTAGTAAAG 904
Db 921 AAKKKKXWAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 980

Qy 905 GGAAGCTGAAGAGTGAACCTGTGAGAAATAGAGCTGTGGAATACACAGTTGTATACCCA 964
Db 981 KKKTKKKKAGKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 1040

Qy 965 ATTCAAGCTGCAAGGGTGTCTGTATGCTGGAGCTTCCTGTCGCTATTCGGGTAAC 1024
Db 1041 KKKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 1100

Qy 1025 ATTTCATAGTGGG 1038
Db 1101 KKKKKKKKKGGGG 1114

```

```

RESULT 12
CNS000DL/c
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR01JL6 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL065414
VERSION AL065414.1 GI:4938827
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of

```

```

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
Pi and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR01JL6"
/clone_lib="RPCI-98"
/notes="end : TET3"

```

ORIGIN

```

Query Match      2.7%; Score 37.8; DB 29; Length 1101;
Best Local Similarity 15.5%; Pred. No. 17;
Matches 51; Conservative 155; Mismatches 122; Indels 1; Gaps 1;

Qy 1049 CTCATTTCATTAAGTCTAACTATTCTGGGTAGGAGTTCTCAATGTCCTCTCTGCTG 1108
Db 1099 MTMHHVYTMTHMYHHTTMMTMMHMTMTMTTTHTTTHMTTTHMTTMTTTCY 1040

Qy 1109 TCACCTTACACATCATGGGGTCAACAGGTATACAGCTTCATAGAGTGGGCATTG 1168
Db 1039 MCHHHHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMH 980

Qy 1169 AGTAGCTACCGCATCAAGCCCGGAGCGGTTCACAGCATGGCGCTACGATACATAG 1228
Db 979 TMTTHAYHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMH 920

Qy 1229 AGTCATAGAAACATAAAGAGGCTTCAAGAACCATTCATCAATCTCAAGGTCCTCTCT 1288
Db 919 WCTATCTTAHAAHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMH 861

Qy 1289 TCTGCTACATCAAGAAATCATACACTCAACAGCAAGTCTTCTTCTATCTTCCTATAGC 1348
Db 860 TMHTTAHYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMY 801

Qy 1349 AATTCCTCAAAACCCATCAATCAACCTAAC 1377
Db 800 HCCHCCMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMHMH 772

```

RESULT 13

```

BX381961/c
LOCUS
DEFINITION BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSODI072YF05 3-PRIME, mRNA sequence.
ACCESSION BX381961
VERSION BX381961.1 GI:30453007
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
Li, W.-B., Gruber, C., Jesse, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
E-mail: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODI072CC03NP1.
Location/Qualifiers
1..1201
FEATURES
source

```

FEATURES source

ORIGIN

	Query Match	2.7%;	Score 37.6;	DB 9;	Length 688;
	Best Local Similarity	54.3%;	Pred. No. 15;		
	Matches 76;	Conservative 0;	Mismatches 64;	Indels 0;	Gaps 0;
QY	1056	TTCAATTGTAATCTAAACTATTCTGGGTAGGAGTTCTCAANGTCTTCTCGCTGTCACTTA	1115		
Db	20	TGCAAGTGAACCCACGAGTGTTCAGAGTAGGAACTGCAAAAGCCCTCTCGCAATGATTCA	79		
QY	1116	CAACATCATCANGGGGTTCACACGATATACAGCTTTCATAGAGAGTGGCGCAATTGAACTAGC	1175		
Db	80	GAACAAATGGGTAGTAAATGAAGTCAAAAAAATCTAACAGCAACTTGGATATAGAAATAGT	139		
QY	1176	TACCGCATCGAACCCCGGAAG	1195		
Db	140	GCTATGTCAACCAATTGGAG	159		

A7960759

SHOCK LOCUS

DEFINITION

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

RESEARCH DESIGN

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

sources


```
/clones="UUGC2M0228L20"
/lab_host="E. coli strain XL10-Gold, T1-resistant, P-"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWP42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
```

ORIGIN

```
Query Match      2.7%; Score 37.4; DB 28; Length 598;
Best Local Similarity 60.2%; Pred. No. 16;
Matches 62; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 1220 GATACATAGAGTCATAGAAACATAAAGGAGCTTGAAGAACCATTCMAATCCTTAAGGTC 1279
      |||||
Db 232 GATATATACATCAAGTGTACCATATAGGAAGTACATAAAGTTCAGTGGTTAGTTTC 291
      |||||

Qy 1280 TCTCTCTTTCTGCATCAGATCATCACTCAAAACCA 1322
      |||||
Db 292 ACACTTTTGTCTCATCATATATTAATCACTCCAAAATCCA 334
      |||||
```

Search completed: June 10, 2004, 13:10:43
Job time : 2712.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 08:08:35 ; Search time 421.164 Seconds

(without alignments)
13899.608 Million cell updates/sec

Title: US-10-070-386-1

Perfect score: 1378

Sequence: 1 gtcagctgggtgggtgatat.....accatcaatcaacacaa 1378

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1378	100.0	1378	5	Aaf79834 Mycelia s
2	48.2	3.5	2000	7	Ada71938 Rice gene
C 3	40	2.9	6197	6	Abn80257 Human che
C 4	37.2	2.7	2000	7	Ada71938 Rice gene
C 5	36.2	2.6	2672	6	Aad29004 Bugula ne
C 6	36.2	2.6	4590	5	Aab24065 Yeast AOD
C 7	35.4	2.6	993	4	Aaf83969 Human SER
C 8	35.4	2.6	993	9	Ade29362 Human cDN
C 9	35	2.5	5971	6	Ab133237 Human imm
C 10	35	2.5	207433	5	Abz72040 Gene 216
C 11	35	2.5	207433	7	Abx74891 BAC109812
C 12	34.6	2.5	5059	2	Abx84332 Stealth v
C 13	34.4	2.5	2799	7	Abv74497 Human typ
C 14	34.4	2.5	8805	6	Abk40016 Human che
C 15	34.4	2.5	32874	8	Ada02648 Human TBX
C 16	34.4	2.5	32874	9	Adh72386 Human TBX
C 17	34.2	2.5	1500	9	Adc92243 Human che
C 18	34	2.5	6283	6	Abk39991 Human che
C 19	34	2.5	6283	6	Ab132834 Human imm
C 20	33.4	2.4	403	4	Aai36496 Probe #51
C 21	33.4	2.4	403	4	Abz26537 Probe #50
C 22	33.4	2.4	1781	2	Aav35130 Mouse WRN
C 23	33.4	2.4	4206	2	Aav35115 Mouse WRN

24	33.4	2.4	4792	2	AAX83004	Aax83004 Mouse WRN
25	33.4	2.4	5058	2	AAV35114	Aav35114 Mouse WRN
26	33.4	2.4	6476	2	AAX24302	Aax24302 Murine mw
27	33.4	2.4	7189	4	AAS30652	Aas30652 DNA encod
28	33.4	2.4	7189	4	AAS28725	Aas28725 Genomic s
29	33.4	2.4	7189	7	ACA03415	ACA03415 DNA encod
30	33.4	2.4	7189	8	ADB96763	Adb96763 Novel lun
31	33.4	2.4	16442	2	AAX83006	Aax83006 Partial m
32	33.2	2.4	891	4	AAK83873	Aak83873 Human imm
33	33.2	2.4	892	4	AAK82975	Aak82975 Human imm
C 34	33.2	2.4	6394	6	ABK31374	Abk31374 Signal tr
C 35	33.2	2.4	6394	6	AAK61278	Aak61278 Human gen
C 36	33.2	2.4	14253	6	ABL33495	Ab133495 Human imm
C 37	33.2	2.4	22214	6	ABL01222	Ab101222 Human AKR
C 38	33.2	2.4	22214	6	ABL01105	Ab101105 Human AKR
C 39	33.2	2.4	34548	6	ABL70603	Ab170603 Chemical
C 40	33	2.4	429	6	ABL64333	Ab164333 Stomach c
C 41	33	2.4	429	6	ABL61677	Ab161677 Colon ade
C 42	33	2.4	478	3	AAK36866	Aak36866 Arabidops
C 43	33	2.4	1189	4	AAK83871	Aak83871 Human imm
C 44	33	2.4	2207	4	ABL24548	Ab124548 Drosophil
45	32.8	2.4	414	8	ADA29990	Ada29990 DNA encod

ALIGNMENTS

RESULT 1

AAAF79834
ID AAF79834 standard; DNA; 1378 BP.
AC AAF79834;
XX
DT 30-MAY-2001 (first entry)
XX
DB Mycelia sterilia promoter sequence.
XX
KW Promoter; terminator; regulatory region; filamentous fungus;
KW protein production; ds.
XX
OS Mycelia sterilia.
FN WO200118219-A1.
PD 15-MAR-2001.
XX
PF 07-SEP-2000; 2000WO-JP006104.
XX
PR 07-SEP-1999; 99JP-00252851.
XX
PA (MEIJ) MEIJI SEIKA KAISHA LTD.
XX
PI Watanabe M, Murakami T;
XX
DR WPI; 2001-235202/24.
XX
PT New promoter and terminator functioning synchronously, useful for
PT regulating expression of endogenous gene in filamentous fungi,
PT particularly transformed Mycelia sterilia capable of producing high
PT yields of target protein/substance.
XX
PS Claim 2; Page 18-20; 26pp; Japanese.
XX
CC The present invention provides the sequences of a promoter and terminator
CC from Mycelia sterilia which can be used to regulate the expression of an
CC endogenous gene in filamentous fungi, particularly transformed Mycelia
CC sterilia, to produce useful target proteins or other substances. The
CC present sequence is the promoter of the invention
SQ Sequence 1378 BP; 396 A; 289 C; 340 G; 353 T; 0 U; 0 Other;

Query Match 100.0%; Score 1378; DB 5; Length 1378;
Best Local Similarity 100.0%; Pred. No. 0;

Best Local Similarity 10.0%; Pred. No. 0.00025; Matches 79; Conservative 339; Mismatches 372; Indels 1; Gaps 1;

Qy 585 CTGAGCAACGCCCGGGTTGTAGACTTATCATGATCCACATCTTAAGCAGGATCTG 644
:
Db 28 SCARMGSSRMKSGMSKYKSCSCCKMTTRRSKWYSASSASGRGSKWSSGSYSG 87
:
Qy 645 AGACATTTCCTAAGGCATCCATATAGCATTTGGGCGCTAAAGTCGCGATGAAGATAA 704
:
Db 88 KGMKKYKKSRRWRGRRGRMRSRNRWGRVRCARSGRMRGSMGWSKMSRYMYM 147
:
Qy 705 GGCGGGTGTAAGAGTGGTGTCAAAGAGGAGGTGCGATTACGAGCGCTAAGCAG 764
:
Db 148 CYARGGSCSKKSKGSGNGKTCTRGARGGSGMSGAKYKSGMSKRWMSGCGS 207
:
Qy 765 GTGGGCTAGCAGCTCTCGCAGCTGTGAATAACGTCACCTGCTTAGGTATGTCCACTAA 824
:
Db 208 RSAYSRYGYTSRKYGTYKWTYYASRCWRAYMTTSYWACSSTYTWCBSKRSMWKMR 267
:
Qy 825 TGTGACGAGATCAAAATGCTGATGGTGTAAAATGGCATGTAGTGTAGGTGCCGAAAC 884
:
Db 268 KMFKSYSHYSHYKMWCTAYKKSYYRWCWTRGGSNCRATYWGRTWMSRMWYK 327
:
Qy 885 ACCTTTAGATCTAGTTAAAGGAGGAGCTGAAAGCTCAACCTGTCAAGAATAAGCCTGTGG 944
:
Db 328 KMTWYRGYKMKRGMWAGRWMSRMCWSKACYWYRWWMRTRRRRWAKSSRTSRKK 387
:
Qy 945 AATACAAGTTGATACCACCAATTCAGTCTGCAAGGTCCTCGATATCTGTCGAGCTCCC 1004
:
Db 388 RKCWKRKYKRWGYSRMSCKRAKRWKRGSRGWKMGRCGCMTCRMSKSYGMRWRSW 447
:
Qy 1005 TGTCGCAATTTGGGGGTAACTATTTTCATAGTGGGCGAGAAATGCAACTCTATTTCATTTGA 1064
:
Db 448 KRMAKYKMSRMYRWRKKCSRTTWGKTGGMWGTWGRCRYKRSQMKRKRWRGR 507
:
Qy 1065 ATCTAAACTATTTCTGGGTAGGAGTCTCAATGGTCTTCTCGTGTCACTTACACATCA 1124
:
Db 508 MYRMRKRYNSARYTMRYCARPKKYSYSAARKARCWYRGKYWAGMMKGYKEMMYKM 567
:
Qy 1125 TGGGGTCAACACGATATACAGCTTCA-TAGAGAGTGGCGGATTAAGTAGCTACCGCAT 1183
:
Db 568 MWYKEKYKCSWYKSWYASOMSKSARAGAKMKCRSKMSAWSKMSRSSKCKASKR 627
:
Qy 1184 CGAACCCGGAAGCGGTTCAAGACATCGGGGTACGTAGTAGTACATGAGTCTAGAAACATA 1243
:
Db 628 SSAKRYANMGENTSGSRMSRWSYTCYWRKWSMKSTCTWYTKSKYTKYKSYNYRY 687
:
Qy 1244 AAAGGAGCTTGAAGACCACTCAAAATCCTAAGGGTCTCTCTCTTCGTGCATCACATCA 1303
:
Db 688 RAWMTWMEYIYRYSINTYNTMTSTRMTAMTKYSGRITWSYKICKSKSKYSRW 747
:
Qy 1304 GAATCATACACTCAACACGAGAACTCTTCTATCTTCCCTATAGCAATTCOCACCA 1363
:
Db 748 YVWSMNAXTKWMMRRYATRMMWYRYSKMWYTNCTWGWYTWYRMYMRYMYKCT 807
:
Qy 1364 TCBAATCAACT 1374
:
Db 808 KTYWWSATYW 818
:

RESULT 3
ABN80257/c ID ABN80257 standard; DNA; 6197 BP.
XX AC ABN80257;
XX DF 15-JUL-2002 (first entry)
XX DE Human chemically modified disease associated gene SEQ ID NO 274.
XX KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
XX KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;

PI Spaderna SK, Shimkets RA, Taupier RJ;
 XX WPI; 2003-615999/58.
 DR P-PSDB; ADE23363.
 XX
 PT New serine/threonine protein-kinase like polypeptides and genes, useful
 PT in gene therapy, diagnosis or prognosis of e.g. diseases of blood
 PT coagulation, obesity; diabetes, or cancers.
 XX
 PS Claim 9; SEQ ID NO 8; 77pp; English.
 XX
 CC The invention relates to a new isolated polypeptide has a sequence
 CC (designated SER1, SER2, SER3, SER4, SER5 and SER6, respectively) fully defined
 CC in the specification for their mature forms, variants or fragments). Also
 CC included are an isolated nucleic acid molecule (comprising: a nucleic
 CC acid sequence encoding any of the SER polypeptides above; a nucleic acid
 CC fragment encoding at least a portion of the SER polypeptides or their
 CC variants; or their complements), vectors comprising the nucleic acid
 CC molecules, cells comprising the vectors, an antibody that binds
 CC immunospecifically to the SER polypeptide, determining the presence or
 CC amount of the SER polypeptide or nucleic acid in a sample, identifying an
 CC agent that binds to the SER polypeptide, identifying a (potential)
 CC therapeutic agent for use in the treatment of a pathology related to
 CC aberrant expression or aberrant physiological interactions of the SER
 CC polypeptide, modulating the activity of the SER polypeptide,
 CC pharmaceutical compositions comprising a carrier (and the SER
 CC polypeptide, nucleic acid molecule or antibody), kits comprising in one
 CC or more containers any of the compositions above, screening for a
 CC modulator of activity or latency or predisposition to a pathology
 CC associated with the SER polypeptide and determining the presence of or
 CC predisposition to a disease associated with altered levels of the SER
 CC polypeptide or nucleic acid molecule. The SER polypeptide, nucleic acid
 CC or anti-SER antibody is useful for treating or preventing a pathological
 CC state or pathology associated with the SER polypeptide in a subject,
 CC particularly a human. These pathologies include diseases of blood
 CC coagulation, human liver, hepatoma cells or hepatocellular cells,
 CC obesity, diabetes, or cancers. The polypeptide or polynucleotide is also
 CC useful in gene therapy, diagnostics, prognosis, or research. The present
 CC sequence encodes SER4.
 XX
 SQ Sequence 993 BP; 229 A; 305 C; 264 G; 195 T; 0 U; 0 Other;
 Query Match 2.6%; Score 35.4; DB 9; Length 993;
 Best Local Similarity 54.1%; Pred. No. 2.6;
 Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
 QY 904 GCGAAGCTGAAAGCTGACCTGTGAGATATACCTGTGGATACAAAGTTGATTAACCC 963
 Db 708 GCGAAGCTCAGGGCTGTACGGTGCATAGAGGGCTGCTGGCGAGGCGATAGATTCTC 649
 QY 964 AATTGAGTCTCAAGGCTGTCTGTATGCTGGAGCTTCCCTGTCCATTGTGGGTAAC 1023
 Db 648 AGTTGTCCCTGGAGGCTGCATACACTCACCAGCGCTCGAAGCCCTCTTGAGGCTGGC 589
 QY 1024 TATTTATAGTGG 1036
 Db 588 CAGCTCGTAGAGG 576
 RESULT 9
 ID ABL33237/c
 XX ABL33237 standard; DNA; 5971 BP.
 AC ABL33237;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 DE Human immune system associated gene SEQ ID NO: 1210.
 XX Human; immune system disease; cytosine methylation; antiaesthatic;
 KW antiaesthatic; antiaesthatic; antiaesthatic; antiaesthatic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;

KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.
 XX Homo sapiens.
 OS
 XX WC200200928-A2.
 PN
 XX 03-JAN-2002.
 PD
 XX 02-JUL-2001; 2001WO-BP007537.
 PF
 XX 30-JUN-2000; 2000DE-01032529.
 PR
 XX 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIC-) EPICENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-130909/17.
 DR
 XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 PT
 XX Claim 1; SEQ ID NO 1210; 32pp + Sequence Listing; German.
 PS
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX
 SQ Sequence 6971 BP; 1665 A; 242 C; 1831 G; 3232 T; 0 U; 1 Other;
 Query Match 2.5%; Score 35; DB 6; Length 6971;
 Best Local Similarity 53.2%; Pred. No. 10;
 Matches 74; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
 QY 1238 AACATATAAAGGAGCTTGAAGAACCAATTCATAAGGCTCTCTCTTTCTGCATCA 1297
 Db 2068 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2009
 QY 1298 CATCAAGATCATACACTCAAAACCAAGAACTTTTCTATCTTCTCTATAGCAATTC 1357
 Db 2008 AAATATAACTATACAAAAATCCCAAAAAAACTTCTAAAGCTCCCTCACTCCACCCCA 1949
 QY 1358 AACCCATCAATCAACTAA 1376
 Db 1948 CAACATCCCCCACTAAA 1930
 RESULT 10
 ID ABZ72040/c
 XX ABZ72040 standard; DNA; 207433 BP.
 AC ABZ72040;
 XX
 XX 03-APR-2003 (first entry)
 DT
 DE Gene 216 H194BAC1098L22 nucleotide sequence SEQ ID NO 5.
 XX Human; Gene 216; chromosome 20p13-p12; antiaesthatic; anorectic;
 KW antiinflammatory; gastrointestinal; gene therapy; vaccine; asthma;
 KW obesity; inflammatory bowel disease; promoter; gene; ss.
 OS
 XX Homo sapiens.
 XX WO200178894-A2.
 PN

XX PD 25-OCT-2001.
XX PF 13-APR-2001; 2001WO-US012245.
XX PR 13-APR-2000; 2000US-00548797.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Keith T;
XX DR WPI; 2001-639428/73.
XX DR P-PSDB; ABR00926.
XX PT isolated genes (Gene 216) from human chromosome 20p13-p12 and the
XX PT proteins they encode, useful for the prevention, diagnosis and treatment
XX PT of asthma, obesity and inflammatory bowel disease.
XX PS Example 4; Fig 7; 520pp; English.
XX CC The invention relates to isolated genes (Gene 216) from human chromosome
XX CC 20p13-p12 and the proteins they encode. The nucleic acids and proteins
XX CC may be used in the prevention, diagnosis and treatment of diseases
XX CC associated with inappropriate Gene 216 expression. For example, the
XX CC nucleic acids (or vectors) and proteins may be used to treat disorders
XX CC associated with decreased expression by rectifying mutations or deletions
XX CC in a patient's genome that affect the activity of Gene 216 by expressing
XX CC inactive proteins or to supplement the patients own production of Gene
XX CC 216 proteins. Additionally, the nucleic acids may be used to produce the
XX CC secreted Gene 216 protein, by inserting the nucleic acids into a host
XX CC cell and culturing the cell to express the protein. The nucleic acids and
XX CC complementary sequences may also be used as DNA probes in diagnostic
XX CC assays to detect and quantitate the presence of similar nucleic acid
XX CC sequences in samples and therefore which patients may be in need of
XX CC restorative therapy. The Gene 216 protein may also be used as antigens in
XX CC the production of antibodies against Gene 216 and in assays to identify
XX CC modulators of Gene 216 expression and activity. The anti-Gene 216
XX CC antibodies and antagonists may also be used to down regulate expression
XX CC and activity. The anti-Gene 216 antibodies may also be used as diagnostic
XX CC agents for detecting the presence of Gene 216 proteins in samples (e.g.
XX CC by enzyme linked immunosorbant assay or ELISA). Disorders that may be
XX CC prevented, diagnosed and/or treated by the above methods include, for
XX CC example asthma, obesity and inflammatory bowel disease. The present
XX CC sequence is that of the Gene 216 genomic nucleic acid sequence, promoter
XX CC or enhancer
XX SQ Sequence 207433 BP; 52775 A; 51290 C; 51698 G; 51670 T; 0 U; 0 Other;
Query Match 2.5%; Score 35; DB 5; Length 207433;
Best Local Similarity 55.3%; Pred. No. 66;
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 1210 GGCCTAGCTAGATACATAGAGTCTATAGAAACATATAAGAGGCTTGAAGAACCAATTCAAAT 1269
Db 1335 GTCTTACTCTTCTTCTTAAACACATACACAAATCCAGGGCTAGTAGGCCCTTGAAT 1276
QY 1270 CCTAAGGGTCTCTTCTTCTTCTGTCATCATCAATCAAGATCATACACTCAACACCGGAATC 1329
Db 1275 ATACAGAACATCTCTTCTTAAAGAAACCTCAGGGATATAAAATCCAAATAGAAACCC 1216
QY 1330 TTT 1332
Db 1215 ATT 1213
RESULT 11
ABX74891/C
ID ABX74891 standard; DNA; 207433 BP.
AC ABX74891;
XX 07-APR-2003 (first entry)
XX

DE BAC1098L22 DNA sequence.
XX Gene 216; antiasthmatic; antiinflammatory; ss; anorectic;
XX chromosome 20p13-p12; single nucleotide polymorphism; SNP; gene therapy;
XX respiratory disease; asthma; obesity; bronchial hyper-responsiveness;
XX chronic obstructive pulmonary disease;
XX adult respiratory distress syndrome; inflammatory bowel syndrome.
XX OS Synthetic.
XX PN WQ200283077-A2.
XX PD 24-OCT-2002.
XX PF 15-APR-2002; 2002WO-US012063.
XX PR 13-APR-2001; 2001US-00834597.
XX PR 13-APR-2001; 2001WO-US012245.
XX PA (SCHE) SCHERING CORP.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Keith T, Little RD, Van Berdewegh P, Dupuis J, Del Mastro RG;
XX Simon J, Allen K, Pandit S;
XX WPI; 2003-092960/08.
XX PT New isolated gene 216 nucleic acids, useful for diagnosing, preventing or
XX PT treating a disorder, such as asthma, bronchial hyper-responsiveness,
XX PT chronic obstructive pulmonary disease, obesity or inflammatory bowel
XX PT syndrome.
XX PS Example 6; Fig 7; 650pp; English.
XX CC This invention relates to a novel isolated nucleic acid, gene 216,
XX CC identified from human chromosome 20p13-p12. The invention also discloses
XX CC regions of the 216 gene that contain single nucleotide polymorphisms
XX CC (SNP's) which may be used as markers for disease susceptibility or
XX CC severity. The nucleotides of the invention may have antiasthmatic,
XX CC antiinflammatory or anorectic activities and may be used in gene therapy.
XX CC The nucleic acids, antibodies or its fragments are useful for diagnosing,
XX CC preventing or treating a disorder, such as respiratory diseases (e.g.
XX CC asthma, bronchial hyper-responsiveness, chronic obstructive pulmonary
XX CC disease or adult respiratory distress syndrome), obesity, or inflammatory
XX CC bowel syndrome. The nucleic acids are also useful for identifying
XX CC increased susceptibility of a subject to the disorders mentioned. The
XX CC nucleic acids can also be used as primers and templates for the
XX CC recombinant production of disorder-associated peptides or polypeptides,
XX CC for chromosome and gene mapping, or for tissue distribution studies. The
XX CC present sequence represents a gene 216 cDNA sequence used in the scope of
XX CC the invention
XX SQ Sequence 207433 BP; 52775 A; 51289 C; 51698 G; 51671 T; 0 U; 0 Other;
Query Match 2.5%; Score 35; DB 7; Length 207433;
Best Local Similarity 55.3%; Pred. No. 66;
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 1210 GGCCTAGCTAGATACATAGAGTCTATAGAAACATATAAGAGGCTTGAAGAACCAATTCAAAT 1269
Db 1335 GTCTTACTCTTCTTCTTAAACACATACACAAATCCAGGGCTAGTAGGCCCTTGAAT 1276
QY 1270 CCTAAGGGTCTCTTCTTCTTCTGTCATCATCAATCAAGATCATACACTCAACACCGGAATC 1329
Db 1275 ATACAGAACATCTCTTCTTAAAGAAACCTCAGGGATATAAAATCCAAATAGAAACCC 1216
QY 1330 TTT 1332
Db 1215 ATT 1213
RESULT 12
AAK84332


```

FT      specification, and is included to maintain the base
FT      numbering given in the specification"
FT      3812
FT      /tag= ab
FT      /note= "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the base
FT      numbering given in the specification"
FT      3813
FT      /tag= ac
FT      /note= "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the base
FT      numbering given in the specification"
FT      3818
FT      /tag= ad
FT      /note= "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the base
FT      numbering given in the specification"
FT      3819
FT      /tag= ae
FT      /note= "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the base
FT      numbering given in the specification"
FT      3824
FT      /tag= af
FT      /note= "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the base
FT      numbering given in the specification"
FT      3827
FT      /tag= ag
FT      /note= "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the base
FT      numbering given in the specification"
FT      3841
FT      /tag= ah
FT      /note= "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the base
FT      numbering given in the specification"
FT      3853
FT      /tag= ai
FT      /note= "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the base
FT      numbering given in the specification"
FT      3857
FT      /tag= aj
FT      /note= "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the base
FT      numbering given in the specification"
FT      3860
FT      /tag= ak
FT      /note= "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the base
FT      numbering given in the specification"
FT      3863
FT      /tag= al
FT      /note= "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the base
FT      numbering given in the specification"

Query Match      2.5%; Score 34.6; DB 2; Length 5059;
Best Local Similarity 15.2%; Pred. No. 12;
Matches 69; Conservative 155; Mismatches 224; Indels 0; Gaps 0;

QY      907 AAGCTGAAGCTGAACCTGTCAGAAATAAGCCTGTGGATACAACTTGATAACCAAT 966
DB      3223 HVRADNASNSTATWATCTNNVRANGNSHHCHWGNRABYCDKSNCTSTAT 3282
QY      967 TCAGTCTCAAGGTCCTGATATGCTGGAGCTCCCTGCTCCATCTGGGGTAATAT 1026
DB      3283 HVRSDNAGRATSNAGARSGSWTHASNYARMATYKGASTSSASMDTHATTHGNCMSSTS 3342
QY      1027 TTCATAGTGGGAGGAATGCACTCTATTTCATTTGAATCTAACTATCTGGGTAGGA 1086
DB      3343 MTRAGNNTSPATHRTANASANNTRNGTHCYTWGAVRAGNWTABHCHWNTNSDNTDHYEM 3402
QY      1087 GTTCTCAATGGTCTTCTCGCTGTCACTTACACATCATGGGGCTCAACAGCTATACAG 1146
      | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |

```

```

DB      3403 GYSARCHNGSNGSNCSNTHDNATRACTDRMCTRSSTATHVRSSSSSSSSSSSTHDSGNA 3462
QY      1147 CTTATAGAGAGTGGCGCATTTGAAGTAGCTACCGCATCGAACCCGGAAGCGTTCAAGAC 1206
DB      3463 TTRRSTRTNSNCDDBYTHNNGNHNCHWVSRRSTTHNSHRTGRNMSTTHNSCSWRBTANDR 3522
QY      1207 ATGGCGGTACGTAGATACATAGAGTCATAGAAACATAAAAGGAGCTTGAAGAACCATTTCA 1266
DB      3523 MTHADSTRANHCWVANDARNTRSNNTWTHADSTRANBTARRSNNTWTHDSTRANHCWVCHM 3582
QY      1267 AATCCTAAGGTCCTCTCTCTCTCTGTCATCATCATCAAGAAATCATACACTCAACACAGGAA 1326
DB      3583 VRTNSTATHVRSCNSWTHMGYNBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBT 3642
QY      1327 CTCCTTCTATCTTCCCTATAGCAATTC 1353
DB      3643 BTBTCTCTBTCTCTCNCNBTTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBT 3669

RESULT 13
ABV74497
ID ABV74497 standard; DNA; 2799 BP.
XX
AC ABV74497;
XX
DT 29-JAN-2003 (first entry)
XX
DE Human type 1 diabetes receptive gene polymorphism 3 SEQ ID NO 7.
XX
KW Human; type 1 diabetes receptive gene; gene; ds.
XX
OS Homo sapiens.
XX
PN JP2002238571-A.
XX
PD 27-AUG-2002.
XX
PF 15-FEB-2001; 2001JP-00038621.
XX
PR 15-FEB-2001; 2001JP-00038621.
XX
PA {UYKY-} UNIV KYUSHU.
XX
DR WPI; 2003-032789/03.
XX
PT Primers for identifying of type 1 diabetes-receptive gene.
XX
PS Claim 7; Page 9-10; 11pp; Japanese.
XX
QC The invention relates to a primer pair used for identifying type 1
CC diabetes receptive gene. The present sequence is that of a polynucleotide
CC expressing a type 1 diabetes receptive gene polymorphism
XX
SQ Sequence 2799 BP; 601 A; 750 C; 880 G; 568 T; 0 U; 0 Other;

Query Match      2.5%; Score 34.4; DB 7; Length 2799;
Best Local Similarity 52.0%; Pred. No. 9.7;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY      875 TGCCGAAACACGTTTAGATCTAGTTAAAGGGAAGCTGAAGCTGAACCTGTGAGAAATA 934
DB      81 TGGGAGAGACTCGGGTGAAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 140
QY      935 AGCCTGTGGATACAAAGCTGTGATTAACCAATTCAGTCGTCAAGGGTGTCTGTATGCT 994
DB      141 TTAAGATATTTTCAAGTACTCTTAAAGATGTTAAATCTTTTATCTATATCTGATATGCA 200
QY      995 GGAGCTTCCCTGCTCGCATTTGGGGTAA 1022
DB      201 AATATTCCATGACACCTTGTGGAGTGA 228

RESULT 14

```

Db	1113	ACCTTTAAATAAATAAATATCTTACCAATATTTTATATAAACAACCGAAACCCGA	105
Qy	1357	AAACCCATCAATCAACCTAA	1376
Db	1053	AAACTACTATAAATAAATAA	1034
RESULT 15			
ADA02648			
ID	ADA02648	standard; DNA; 32874 BP.	
XX	AC	ADA02648;	
XX	DT	06-NOV-2003 (first entry)	
XX	XX	Human TBX21 carcinoma associated gene, SEQ ID NO:1166.	
XX	XX	Human; carcinoma associated; oncogene; carcinoma; cancer; breast;	
KW	KW	prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;	
KW	KW	gene, ds.	
XX	OS	Homo sapiens.	
XX	XX	WO2003057146-A2.	
PN	XX	17-JUL-2003.	
XX	PD	26-DEC-2002; 2002WO-US041414.	
XX	PF	26-DEC-2001; 2001US-00035832.	
PR	XX	(SAGR-) SAGRES DISCOVERY.	
XX	PA	Morris DW;	
XX	PI	WPI; 2003-587068/55.	
XX	DR	New recombinant nucleic acid encoding carcinoma associated protein,	
XX	PT	useful for preparing compositions for treating carcinomas.	
XX	PT	Claim 1; SEQ ID NO 1166; 245pp; English.	
PS	XX	The invention relates to recombinant carcinoma associated (CA) nucleic	
XX	CC	acid sequences from mouse and human (ADA01482-ADA03094), and to	
XX	CC	recombinant carcinoma associated proteins (CAP) encoded by them. The	
XX	CC	invention also encompasses expression vectors and host cells comprising a	
XX	CC	CA nucleic acid, a polypeptide (especially an antibody) that specifically	
XX	CC	binds to the protein, and a biochip comprising CA nucleic acid or	
XX	CC	fragments thereof. The sequences of the invention were identified using	
XX	CC	oncogenic retroviruses, which insert into the genome of the host organism	
XX	CC	at random. Many of these do not carry transduced host oncogenes or	
XX	CC	pathogenic trans-acting viral genes, meaning that cancer incidence is a	
XX	CC	direct consequence of the effects of proviral integration into host	
XX	CC	proteocarcinomas. The CA nucleic acid sequences can be used to diagnose	
XX	CC	carcinoma (especially breast cancer, prostate cancer, lymphoma or	
XX	CC	leukaemia) or a propensity to carcinoma by determination of the sequence	
XX	CC	of a CA gene, or by determination of CA gene expression in particular	
XX	CC	tissues. CA nucleic acids, proteins and antibodies are also useful as	
XX	CC	therapeutic agents and in screening and evaluating drug candidates. The	
XX	CC	present sequence represents a specifically claimed human CA nucleic acid	
XX	CC	sequence of the invention. Note: The complete sequence data for this	
XX	CC	patent did not form part of the printed specification, but was obtained	
XX	CC	in electronic format directly from WIPO at	
XX	CC	ftp.wipo.int/pub/published_pct_sequences	
XX	XX	Sequence 32874 BP; 7455 A; 8048 C; 8163 G; 9208 T; 0 U; 0 Other;	
Query Match		2.5%; Score 34.4; DB 8; Length 32874;	
Best Local Similarity		52.0%; Pred. No. 38;	
Matches		77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;	
Qy	875	TGCCGAAACACGTTTAGATCTAGTAAAGGAGCTGAAGCTGACCTTCAGAAATA	934

	Query Match	2.5%	Score 34.4	DB 8	Length 32874
	Best Local Similarity	52.0%	Pred. No. 38		
	Matches	77	Conservative 0	Mismatches 71	Indels 0
1237	AAACATAAAGGAGGCTTGAAGAACATTCAAAATCCTAAGGGTCTCTCTCTTTCTGCATC	1296			
QY					
1173	AAACAACGAAAAATTTCTAAACACCAATAAAGCCCAATAAACTTACCACTTTTAACTC	1114			
DB					
1297	ACATCAAGAAATCATACACTCAAAACGAGAACTCTTTCTATCTTCCCTATAGCAATTC	1356			
QY					

Db 8489 TGGGAGGACTGGGGTGNAGGTAGAGAGAGGAGGAGTTAACTTTAATGACAAAATA 8548
Qy 935 AGCCTGTGGATACAAAGTTGATAACCCAAATTCAGTCGTCAAGGGTGTCTGTATGCT 994
Db 8549 TTAAGTATTTTTCAAAGTACTCTTAAAAATGTAAATCTTTATCTATATATCTGATATGCA 8608
Qy 995 GGACCTTCCCTGTCCCAATGTGGGTAA 1022
Db 8609 AATATTCATGACACCTTGGGATGA 8636

Search completed: June 10, 2004, 09:56:02
Job time : 424.164 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 08:11:45 ; Search time 3792.71 Seconds
(without alignments)
15747.756 Million cell updates/sec

Title: US-10-070-386-1
Perfect score: 1378
Sequence: 1 ggcagctgggtgatat.....accatcaatcaacacaaac 1378

Scoring table: IDENTIFY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues 6940544
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Result No.	Score	Query Match	Length	DB	ID	Description
1	1378	100.0	1378	6	BD013073	BD013073 Regulator
2	48.2	3.5	2000	6	AX655393	AX655393 Sequence
3	42.2	3.1	7218	6	I66494	I66494 Sequence 14
4	41.8	3.0	231585	2	AC103186	AC103186 Rattus no
5	41.8	3.0	329549	2	AC117331	AC117331 Rattus no
6	41.6	3.0	164377	2	AC136809	AC136809 Rattus no
7	41.6	3.0	226170	2	AC105687	AC105687 Rattus no
8	40	2.9	6197	6	AX344849	AX344849 Sequence
9	39.8	2.9	221649	2	AC137392	AC137392 Rattus no
10	39.8	2.9	262254	2	AC103016	AC103016 Rattus no
11	39.8	2.9	342084	2	AC128752	AC128752 Rattus no
12	39.6	2.9	81624	3	AC099763	AC099763 Caenorhab
13	39.6	2.9	290066	2	AC129421	AC129421 Rattus no
14	39.4	2.9	81921	10	AL805938	AL805938 Mouse DNA
15	39.4	2.9	186858	10	AC130210	AC130210 Mus muscu
16	39.4	2.9	195411	10	AL669958	AL669958 Mouse DNA
17	39.4	2.9	218807	2	BX511249	BX511249 Mus muscu
18	39	2.8	237119	10	AC115291	AC115291 Mus muscu
19	38.8	2.8	16512	3	CE049795	U49795 Caenorhab
20	38.8	2.8	34919	3	AF100663	AF100663 Caenorhab
21	38.6	2.8	192211	2	AC114644	AC114644 Mus muscu
22	38.6	2.8	129489	10	AL713860	AL713860 Mouse DNA
23	38.6	2.8	143148	10	AL607034	AL607034 Mouse DNA
24	38.6	2.8	233819	2	AC095098	AC095098 Rattus no
25	38.6	2.8	234873	2	AC137028	AC137028 Rattus no
26	38.6	2.8	240425	10	CNS07YOT	AL713839 Mus muscu
27	38.6	2.8	272563	2	AC095469	AC095469 Rattus no
28	38.4	2.8	195611	10	AC121971	AC121971 Mus muscu
29	38.4	2.8	334371	2	AC109571	AC109571 Rattus no
30	38.2	2.8	149092	2	AC127682	AC127682 Mus muscu
31	38.2	2.8	180907	9	AC131157	AC131157 Homo sapi
32	38.2	2.8	202275	2	AC102341	AC102341 Mus muscu
33	38.2	2.8	253768	2	AC125994	AC125994 Rattus no
34	38.2	2.8	302921	2	AC090998	AC090998 Homo sapi
35	38	2.8	75342	2	AC145944	Continuation (4 of
36	38	2.8	194370	2	AL645472	AL645472 Mus muscu
37	38	2.8	220638	10	AL645723	AL645723 Mouse DNA
38	38	2.8	239946	2	BX663615	BX663615 Danio rer
39	38	2.8	257760	2	AC126638	AC126638 Rattus no
40	37.8	2.7	114999	9	AF003421	AF003421 Homo sapi
41	37.8	2.7	142148	9	AC025465	AC025465 Homo sapi
42	37.8	2.7	167912	9	AC060232	AC060232 Homo sapi
43	37.8	2.7	170066	9	AL161730	AL161730 Human DNA
44	37.8	2.7	180001	9	AC112196	AC112196 Homo sapi
45	37.8	2.7	205224	10	AL671908	AL671908 Mouse DNA

ALIGNMENTS

RESULT 1	BD013073	Regulatory sequences and expression system functional in mold	1378 bp	DNA	linear	PAT 02-AUG-2002
LOCUS	BD013073	Regulatory sequences and expression system functional in mold				
DEFINITION	BD013073	Regulatory sequences and expression system functional in mold				
ACCESSION	BD013073.1	GI:22093262				
VERSION	BD013073.1	GI:22093262				
KEYWORDS	WO 0118219-A/1.					
SOURCE	unidentified					
ORGANISM	unidentified					
REFERENCE	1 (bases 1 to 1378)					
AUTHORS	Watanabe,M. and Murakami,T.					
TITLE	Regulatory sequences and expression system functional in mold fungi					
JOURNAL	Patent: WO 0118219-A 1 15-MAR-2001;					

Pred. No. is the number of results predicted by chance to have a


```
Db 28 SCARMSSRMKMSKYSKYKCSGCKKMTTRKSKWYSASSASGRTGSKMSGSG 87
Qy 645 AGACATTTCTTAAGGCATCCATATAGGCATTTGGCGCTAAGTCGGCATTTGAAGGATAA 704
Db 88 KGMKRYKRSKRWGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGR 147
Qy 705 GGGGGGTGTGAAGTGTGTGTGTCAAAAGGAGGTGTGATGCGTATACACGCGCTTAAGCAG 764
Db 148 CYARGCGSCRRKSKGSGWGTCTRGARGGSGSSGAKYKSGSKRMWMSGGRSGCR 207
Qy 765 GTGGGCTAGCAGCTGTCTGAGCTGTGAATAAGCTCACCTTCCTTAGGTATGTCACCTAA 824
Db 208 RSAYRYGTGRKTYGTGKTYTYSASRCRAYMTISWACSSYTWCRSKRSHWQWR 267
Qy 825 TGTGAGCAGATGCAAAATGCTGATTTGGGTTAAATGCGCATGTAGTGTAGTGTGCGAAGAAC 884
Db 268 KMRWSRGYMSYMSYKMMCTAYKYSYVRCWYRGGRGATRYWGRGYMSRMWYK 327
Qy 885 ACGTTTAGATCTAGTTAAAGGAGCTGAAAGCTGAACCTGTGAGATAAGCTGTGG 944
Db 328 KTYWYRGYKMKRWAGRMMSKRWKSKACYYTMRWMTTRRWKSKSRTSRKK 387
Qy 945 AATACAAAGTTGATAACCAATTCAGTCGTCAGGGGTGTCTGATATGCTGGAGCTTCCC 1004
Db 388 RACWCRKRYKRMGRYSRMRCKEARWKRCSGRAMKGCRCMTCKMSYGMWKS 447
Qy 1005 TGTGCGATTTGGGGTAACTATTATAGTGTGGGCGAAGATGCAACTCTATTTCATTTGA 1064
Db 448 KRMASKYKMSRMVYRKKCSRTTMMGKTGGMGTGRCRYKRGSGMRKCRWRGR 507
Qy 1065 ATCTAAACTATTGCGGTAGAGTCTCAATGGTCTTCTGCTGCTCACTTACACATCA 1124
Db 508 MYRWRKYRYSRYTMYRCARKEYSYSAKARCTYRGKYWAGWMMRYKMYKM 567
Qy 1125 TGGGGGTCAACAGCTATACAGTTCATAGAGTGTGGGCGAAGTGAAGTACCGCAT 1183
Db 568 MMYTKYKCSWYTCMSYASCMKSAKAKCKESKMSAKSKMSRKRCKKASKR 627
Qy 1184 CGAACCGGAGCGTTCAAGACATGGCGTACATAGATAGATAGATAGATAGACATA 1243
Db 628 SSARYAMGGMGTSGSRMWSKSYTCYWRWGSWKSCTWYMSKTYAKYSYRY 687
Qy 1244 AAAGGAGCTTGAAGAACCATCAATCAAGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1303
Db 688 RACWCMYRMYRYRYSYMTYMAWYTSSTRAMTGMKYSGRYMTSWYKCKMSYRSMW 747
Qy 1304 GAATCATACACTCAACAGCAACTCTTTCTAATCTCTCTCTCTCTCTCTCTCTCTCTCT 1363
Db 748 YYWSWMAKTKMWRATRMWMMYRYSMKWYTWCTMGGYTWYMTYMYKMYMYKCT 807
Qy 1364 TCAATCAACT 1374
Db 808 KTYWYNSATYW 818

RESULT 3
LOCUS 166494/c
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 7218)
AUTHORS Dörner, P., Scheiflinger, P. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
source Location/Qualifiers
1. 7218
/organism="unknown"
```

```
ORIGIN
Query Match 3.1%; Score 42.2; DB 6; Length 7218;
Best Local Similarity 8.9%; Pred. No. 0.48; Indels 0; Gaps 0;
Matches 41; Conservative 212; Mismatches 210;
/mol_type="unassigned DNA"

Qy 607 TAGACCTACATGATGACACATCTTAAAGCAGGAGATCTGAGACATTTTCTAAGCATCAT 666
Db 1450 TAGAGAAATTGGTACRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1391
Qy 667 ATAGCATTTGGCGCTAAGTCGGCATTTGAAGAGATGAAGGGGCTGTGAAGTGTGTGT 726
Db 1390 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1331
Qy 727 CAAAGAGGTCGATTTGGCTATACACGCGCTAAGCAGGTCGGGCTAGCAGCTGTCTGCG 786
Db 1330 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1271
Qy 787 CTGTGATAAAGCTCACCTTCTAGGTATGTCACCTAATGTGACGAGATGCAATGCTGA 846
Db 1270 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1211
Qy 847 TTGGGTTAAATGGGCGATGTAGTGTAGGTGCGCAAAACACGTTTATAGTCTAGTAAAGG 906
Db 1210 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1151
Qy 907 AGCTGAAAGCTGCACTGTGAGATAAGCTGTGGAATACCAATGATGATACCAAT 966
Db 1150 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1091
Qy 967 TCAGTCTCAAGGCTGCTGATATGCTGAGGCTTCCCTGCTGCTGCTGCTGCTGCTGCT 1026
Db 1090 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1031
Qy 1027 TCATAGTGGGCGAGATGCAATCTATTTCAATGATCTA 1069
Db 1030 TTAATTTCTGAGCGTATGCGCAACGAGAGAAAATAGTTATA 988

RESULT 4
LOCUS AC103186
DEFINITION Rattus norvegicus clone CH230-107D17, *** SEQUENCING IN PROGRESS
ACCESSION AC103186
VERSION AC103186.5 GI:30578507
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM
Mammalia; Eutheria; Rodentia; Chordata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa;
Rattus.
REFERENCE
1 (bases 1 to 231585)
AUTHORS Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Aisbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, P.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Devila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Gatta, R., Garcia, A., Garner, T., Garza, M.,
Georgiadis, E., Geer, K., Gill, R., Grady, M., Guerr, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
```



```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 230369: contig of 230369 bp in length
* 230370 230469: gap of unknown length
* 230470 231585: contig of 1116 bp in length.
FEATURES             source
    location/Qualifiers
        1..231585
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-107D17"
misc_feature         1..2354
                    /note="wgs_end_extension
                    clone_end:Sp6"
                    5337..6318
misc_feature         /note="clone boundary
                    clone_end:Sp6
                    site:EcORI
                    end_sequence:BH359310"
misc_feature         complement(227205..227963)
                    /note="clone boundary
                    clone_end:T7
                    site:EcORI
                    end_sequence:BH359309"
ORIGIN
Query Match          3.0%; Score 41.8; DB 2; Length 231585;
Best Local Similarity 47.2; Pred.No. 0.71;
Matches 127; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
Qy 1110 CACTTACACACATCATCGGGGTCAACAACGTATACAGCTTCATAGAGAGTGGCGATTGA 1169
Db 168436 CACACACACACACACACACACACACACACACAGAGAGAGAGAGAGAGAGAGAGA 168495
Qy 1170 AGTAGTACCGCATCGAACCCGGGTTCAAGACATGGGCTAGATCGTAGATCATAGA 1229
Db 168496 CAGAGACAGAGAGAGACAGAGACAGACAGAGAGAGACAGACAGAGAGACAGAGA 168555
Qy 1230 GTCATGAAACATAAAGGAGCTTGAAGAACCATTCAAATCCTAAGGTCCTCTCTTT 1289
Db 168556 GGCAAGAGACAGAGAGAGAGAGCAAAATCTAAATATCTCTAGGGAAAACTCTATT 168615
Qy 1290 CTGCATCACATCAAGATCATACACTCAAAACCCAGGAACCTCTTTCTATCTCCCTATAGCA 1349
Db 168616 CTGAGGAACATTTTAAAAACATCATCACACAGTAAATTAACCAATCCAGAAATCTATA 168675
Qy 1350 ATTCCCAAAACCCATCAATCAACCTAACA 1378
Db 168676 CTTTCAATAAACTTTCACTCACTTAAAA 168704

RESULT 5
AC117331
LOCUS          329549 bp      DNA      linear      HTG 09-OCT-2002
DEFINITION    Rattus norvegicus clone CH230-242A22, *** SEQUENCING IN PROGRESS
              ***, 4 unordered pieces.
ACCESSION     AC117331
VERSION       AC117331.4 GI:23194666
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE        Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
REFERENCE     1 (bases 1 to 329549)

```

```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 230369: contig of 230369 bp in length
* 230370 230469: gap of unknown length
* 230470 231585: contig of 1116 bp in length.
FEATURES             source
    location/Qualifiers
        1..231585
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-107D17"
misc_feature         1..2354
                    /note="wgs_end_extension
                    clone_end:Sp6"
                    5337..6318
misc_feature         /note="clone boundary
                    clone_end:Sp6
                    site:EcORI
                    end_sequence:BH359310"
misc_feature         complement(227205..227963)
                    /note="clone boundary
                    clone_end:T7
                    site:EcORI
                    end_sequence:BH359309"
ORIGIN
Query Match          3.0%; Score 41.8; DB 2; Length 231585;
Best Local Similarity 47.2; Pred.No. 0.71;
Matches 127; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
Qy 1110 CACTTACACACATCATCGGGGTCAACAACGTATACAGCTTCATAGAGAGTGGCGATTGA 1169
Db 168436 CACACACACACACACACACACACACACACACAGAGAGAGAGAGAGAGAGAGAGA 168495
Qy 1170 AGTAGTACCGCATCGAACCCGGGTTCAAGACATGGGCTAGATCGTAGATCATAGA 1229
Db 168496 CAGAGACAGAGAGACAGAGACAGAGACAGAGACAGAGACAGAGACAGAGACAGAGA 168555
Qy 1230 GTCATGAAACATAAAGGAGCTTGAAGAACCATTCAAATCCCTAAGGTCCTCTCTTT 1289
Db 168556 GGCAAGAGACAGAGAGAGAGACGAAAATCTAAATATCTCTAGGGAAAACTCTATT 168615
Qy 1290 CTGCATCACATCAAGATCATACACTCAAAACCCAGGAACCTCTTTCTATCTCCCTATAGCA 1349
Db 168616 CTGAGGAACATTTTAAAAACATCATCACACAGTAAATTAACCAATCCAGAAATCTATA 168675
Qy 1350 ATTCCCAAAACCCATCAATCAACCTAACA 1378
Db 168676 CTTTCAATAAACTTTCACTCACTTAAAA 168704

RESULT 5
AC117331
LOCUS          329549 bp      DNA      linear      HTG 09-OCT-2002
DEFINITION    Rattus norvegicus clone CH230-242A22, *** SEQUENCING IN PROGRESS
              ***, 4 unordered pieces.
ACCESSION     AC117331
VERSION       AC117331.4 GI:23194666
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE        Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
REFERENCE     1 (bases 1 to 329549)

```

AUTHORS

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, P., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Centz, A., Chacko, J., Chavez, D., Chen, G., Chen, X., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyie, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gregorogis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huly, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, J., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuwha, L., Louised, H., Lozada, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meener, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, D., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokeleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L., Puzos, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojase, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, D., Waldron, L., Walker, B., Wang, J., Valias, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiser, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.

TITLE

Unpublished
2 (bases 1 to 329549)

JOURNAL

Worley, K. C.
Direct Submission

AUTHORS

Submitted (10-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

JOURNAL

3 (bases 1 to 329549)
Rat Genome Sequencing Consortium.

AUTHORS

Submitted (09-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Sep 19, 2002 this sequence version replaced gi:21746236. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GUAC

Center clone name: CH230-242A22

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 198199 bases at least Q40

Consensus quality: 201002 bases at least Q30

Consensus quality: 202294 bases at least Q20

Estimated insert size: 210461; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 38119: contig of 38119 bp in length
* 38120 38219: gap of unknown length
* 38220 291497: contig of 253278 bp in length
* 291498 291597: gap of unknown length
* 291598 294124: contig of 2527 bp in length
* 294125 294224: gap of unknown length
* 294225 329549: contig of 35325 bp in length.

FEATURES

source

1..329549
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-242A22"
complement(2196..2863)
/note="clone_boundary
notes end:Sp6
site:MboI
and sequence:RXAAF11TV"
38220..39946
/note="wgs_contig"

misc_feature

misc_feature

ORIGIN

Query Match 3.0%; Score 41.8; DB 2; Length 329549;
Best Local Similarity 47.2%; Pred. NO. 0.72;
Matches 127; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 1110 CACTTACACATCATGCGGGTCAACAGGTATACAGCTTCATAGAGTGGCGCATTCGA 1169
DB 278081 CAGAGA 278140
QY 1170 AGTAGCTACCGCATCGAACCCGGAAGCGGTTCAAGACATCGGCGGTACGTAGATACATAGA 1229
DB 278141 CAGAGAGACAGAGAGACAGAGAGACAGAGAGACAGAGACAGAGACAGAGACAGAGA 278200
QY 1230 GTCATAGAACATTAAGGAGGCTTGNAGACCATCTCAATCTTCTCTCTTT 1289
DB 278201 GCGAGAGACAGA 278260
QY 1290 CTGCATCATCAAGAATCATACATCTCAAAACAGGAGACTTTTCTATCTTCCCTATPAGCA 1349
DB 278261 CTGAGGACATTTTAAAAAACATCATCACACAGTAAATTAACCAATCCAGAAATCTATA 278320
QY 1350 ATTCCCAAAACCCCATCATCACTTCAACCTTACCA 1378

Db 278321 CTTTCAATAAAGCTTCACTCACTTAAAA 278349

RESULT 6
AC136809
LOCUS
DEFINITION AC136809 164377 bp DNA linear HTG 23-NOV-2002
***, 8 unordered pieces.
ACCESSION
VERSION AC136809.2 GI:25188374
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Altschrocks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, S., Davy-Carroli, L., De Anda, C., Dederich, D., Deigado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Harvey, R., Havelak, P., Haves, A., Henderson, N., Hernandez, P., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Garbogeorgis, E., Geer, K., Gill, R., Grady, N., Guerra, W., Guevara, W., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huiyuk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Liu, J., Kwis, C., Kraft, C.L., Lebowitz, H., Levan, J., Lewis, J., Li, Z., Liu, J., Liu, J., Liu, W., Lu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Louised, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, N., Mahindartne, M., Mahmoud, M., Mailloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reich, R., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartabeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczky, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 164377)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (23-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 23, 2002 this sequence version replaced gi:24796682.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KCSF
Center clone name: CH230-349D12
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 119076 bases at least Q40
Consensus quality: 123584 bases at least Q30
Consensus quality: 126629 bases at least Q20
Estimated insert size: 127842; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 3175: contig of 3175 bp in length
* 3176 3275: gap of unknown length
* 3276 115030: contig of 111755 bp in length
* 115031 115130: gap of unknown length
* 115131 119148: contig of 4018 bp in length
* 119149 119248: gap of unknown length
* 119249 155611: contig of 36363 bp in length
* 155612 155711: gap of unknown length
* 155712 159726: contig of 4015 bp in length
* 159727 159826: gap of unknown length
* 159827 161136: contig of 1310 bp in length
* 161137 161236: gap of unknown length
* 161237 162452: contig of 1216 bp in length
* 162453 162552: gap of unknown length
* 162453 164377: contig of 1825 bp in length.
* 162553 164377: contig of 1825 bp in length.
----- Location/Qualifiers
1. 164377
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-349D12"
misc_feature 1..1297
note="wgs contig"
misc_feature 1711..3175
note="wgs contig"
misc_feature 3276..4359
note="wgs contig"

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackeleah, O., Okwundu, G., Olaripunsagoun, A., Pal, S., Parks, X., Pasternak, S., Paul, H., Perez, A., Perez, L., Pflannkuch, C., Plogger, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shivartsbeyn, A., Sison, I., Sitter, C.D., Smajls, D., Snead, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Vallas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleciyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 226170)
Worley, K.C.

Direct Submission
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 226170)

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 18, 2002 this sequence version replaced gi:21743897.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

Center project name: GNCC
Center clone name: CH230-1111
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 198186 bases at least Q40
Consensus quality: 202314 bases at least Q30
Consensus quality: 204704 bases at least Q20
Estimated insert size: 229288; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 7584: contig of 7584 bp in length

* 7585 7684: gap of unknown length
* 7685 224621: contig of 216937 bp in length
* 224622 224721: gap of unknown length
* 224722 226170: contig of 1449 bp in length.

FEATURES
source Location/Qualifiers
1. .226170: contig of 1449 bp in length.
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-1111"
misc_feature 45780..46902
/note="wgs_contig"

ORIGIN
Query Match 3.0%; Score 41.6; DB 2; Length 226170;
Best Local Similarity 49.1%; Pred. No. 0.82;
Matches 110; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 1155 AGAGTGGCGCATTTAGTAGTACCGCATCGAACCGGAGCGTTCAAGCATCGGGCT 1214
DB 61562 AGATGCTGGATTAGAGATGACAAAGAGAGGGGGAGACATATGAGAGTCTATCAGG 61503
QY 1215 ACCTAGATACATAGATGATAGAAATAGAGAGCTTTGAAGAACCTTCAAAATCTTAA 1274
DB 61502 ACATGTATGCATATATTACTAGTACGTTCAAGAGCATCACTTCTATTAAATATTATA 61443
QY 1275 GGGTCTCTCTTCTCTCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1334
DB 61442 CATTATCATGTTTTTCTTAAATAAATAAATAAAGTTAAGATACAGAGAACTCTATGA 61383
QY 1335 ATCTTCCCTATAGCAATTCCTCCAAACCCATCAATCAACCTACCA 1378
DB 61382 GACTGAACCTAGAGAAATCCCAACCCCTCTAGATACATACCA 61339

RESULT 8
AX344849/c
LOCUS
DEFINITION Sequence 274 from Patent WO0200927.
ACCESSION AX344849
VERSION AX344849.1 GI:18492735
KEYWORDS
SOURCE synthetic construct
ORGANISM artificial sequences.

REFERENCE
1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with development genes
JOURNAL Patent: WO 0200927-A 274 03-JAN-2002;
EpiGenomics AG (DE)
FEATURES
source Location/Qualifiers
1. .6197
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN
Query Match 2.9%; Score 40; DB 6; Length 6197;
Best Local Similarity 51.7%; Pred. No. 2.3;
Matches 91; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 1200 TCAAGCATGGCGTACGTAGATACATAGATCATAGATACATAGATACATAGATACATAGAT 1259
DB 3228 TAAATAAATAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 3169
QY 1260 CCAATTCAAATCTTAAGGCT 1319
DB 3168 CAATAAATATCTTAATCT 3109
QY 1320 CAGGAACTCTTCTATCT 1375
DB 3108 CAATAAATCTTAATCT 3053

AC137392 221649 bp DNA linear HTG 20-NOV-2002
Rattus norvegicus clone CH230-unknown, *** SEQUENCING IN PROGRESS
***, 5 unordered pieces.
AC137392
AC137392.1 GI:25138502
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (Bases 1 to 221649)
Munzy, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,
Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Aryalabechi, V., Ayvagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, P.,
Biswal, K., Blair, O., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, K., Buhay, C., Burch, P., Burrell, K., Calderon, S.,
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dink, H., Divya, K.,
Draper, H., Dugan-rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, I., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, M., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hughes, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idiebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuhewa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Mailoy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, K., Martinez, E.,
Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokeleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, S., Pul, L.,
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaj, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, K., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejo, Z., Umami, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (Bases 1 to 221649)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center Project name: KZNG

----- Summary Statistics
Center clone name: CH230-unknown

Assembly program: Phrap; version 0.990329
Consensus quality: 150999 bases at least Q40
Consensus quality: 156151 bases at least Q30
Consensus quality: 159735 bases at least Q20

Estimated insert size: 155933; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 155377: contig of 155377 bp in length
* 155378 155477: gap of unknown length
* 155478 217856: contig of 62379 bp in length
* 217857 217956: gap of unknown length
* 217957 218983: contig of 1027 bp in length
* 218984 219083: gap of unknown length
* 219084 220305: contig of 1222 bp in length
* 220306 220405: gap of unknown length
* 220406 221649: contig of 1244 bp in length.

FEATURES

source

1. 221649
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-unknown"

misc_feature

1. 221649
/notes="wgs contig"

misc_feature

32902..34790
/notes="wgs contig"

misc_feature

69822..72586
/notes="wgs contig"

misc_feature

89976..91515
/notes="wgs contig"

misc_feature

155478..156627
/notes="wgs contig"

misc_feature

172910..174635
/notes="wgs contig"

misc_feature

197144..200776
/notes="wgs contig"

ORIGIN

Query Match 2.9%; Score 39.8; DB 2; Length 221649;
Best Local Similarity 50.8%; Pred. No. 2.9;
Matches 95; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 11 GCGTGATATCATCGCTGGTGTCCAAACACTGTTAGTACTACGAATGGAAGAA 70
Db 36230 GTGTTGGATATCTCTCTTTGTAGAAATATGTTAAAGAGACATGAAAAAGGTCAGAG 36289
QY 71 ACGGTGGTGTCTGTGGCAGCTGAAGACTGAAGAGAGCCAAAGATATTCACAATCCGAT 130
Db 36290 AAGGCTTACAGAAGGCTTCTACAAAAGGAAAGCTTTCAACAAGAGAGAAAAGCTGA 36349
QY 131 ACGTTGGATCATATGCTTGTTCAGAGAGACAGCTGCATCTACCTGCTGTTCCCTCTTC 190
Db 36350 AAGCTGGAATCAAAACTGTTCTCCACCTCCCTGTTGTTGTCATCTGCTCTCTCT 36409
QY 191 GTTGTAC 197
Db 36410 GTTGGAC 36416

RESULT 10

AC103016

LOCUS

DEFINITION

AC103016

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC103016 262254 bp DNA linear HTG 13-MAY-2003
Rattus norvegicus clone CH230-171119, *** SEQUENCING IN PROGRESS
*** 5 unordered pieces.

AC103016 GI:30580739

HTG: HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 262254)

Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Ayodeji,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,I., Garza,M.,
Gebregeorgis,B., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,M., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Loulasegh,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,B.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwankweli,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Resier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

FEATURES

Location/Qualifiers
1..81624
/organism="Caenorhabditis briggsae"
/mol_type="genomic DNA"
/strain="Gujarat G16"
/db_xref="taxon:6238"

ORIGIN

Query Match 2.9%; Score 39.6; DB 3; Length 81624;
Best Local Similarity 58.5%; Pred. No. 3.2;
Matches 69; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 34 TCANAACTGTGTAGTACTAGCAATGAGGAAGAACCGTGTGTGGCAGCTGAA 93
|||||
Db 8467 TCCAAAATATCTATTATTATTCGCGAGCTGTAAAAAAGAAAGTTTGTACCAATTGAG 8408
|||||

QY 94 GACTGAGAGGAGCCAGACATATTCACATCGTATCGGTGTCATCAATGCTTCT 151
|||||
Db 8407 ACATGAAATACGTATTAATTAATATCAAGTAAAGCAAAAGTACATGACTGTGATT 8350
|||||

RESULT 13

AC129421/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-62C13, *** SEQUENCING IN PROGRESS
***, 5 unordered pieces.

ACCESSION

AC129421.3 GI:25138103

VERSION

HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

Rattus norvegicus

ORGANISM

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 290066)
Muzny, D., Marie, H., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gragegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hu, Y.K., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loresukewa, L., Loulseg, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, B., Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeleneh, O., Okunolu, G., Oiarupunsagoon, A., Pal, S., Parks, K.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Paaternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, P., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Rellay, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Shen, H., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Keinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 290066)
Worley, K.C.

Direct Submission
Submitted (30-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 290066)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:23267370.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GURV
Center clone name: CH230-62C13
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 216596 bases at least Q40
Consensus quality: 216555 bases at least Q30
Consensus quality: 221692 bases at least Q20
Estimated insert size: 216915; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1 40609: contig of 40609 bp in length
* 40610 40709: gap of unknown length
* 40710 281950: contig of 241241 bp in length

AC130210/c	AC130210	186958 bp	DNA	linear	ROD 27-NOV-2003
LOCUS	Mus musculus BAC clone RP24-344N22	from chromosome 7,	complete		
DEFINITION	sequence.				
ACCESSION	AC130210				
VERSION	AC130210.4	GI:30231364			
KEYWORDS	HTG.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 186958) Nguyen,C. and Kozlowicz,A.				
TITLE	The sequence of Mus musculus BAC clone RP24-344N22				
JOURNAL	Unpublished (2001)				
REFERENCE	2 (bases 1 to 186958)				
AUTHORS	Wilson,R.				
TITLE	Sequencing of Mus musculus				
JOURNAL	Unpublished (2001)				
REFERENCE	3 (bases 1 to 186958)				
AUTHORS	McPherson,J.D. and Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-AUG-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE	4 (bases 1 to 186958)				
AUTHORS	McPherson,J.D. and Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (03-APR-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE	5 (bases 1 to 186958)				
AUTHORS	McPherson,J.D. and Waterston,R.H.				

```

/rpt_family="MERR2_type
9419..9580
/rpt_family="B4"
10335..10537
/rpt_family="B2"
11210..11361
/rpt_family="B4"
13411..11497
/rpt_family="Alu"
11810..11962
/rpt_family="B4"
12046..12433
/rpt_family="ERVK"
12680..12727
/rpt_family="ERV1"
12774..12972
/rpt_family="B2"
13389..13502
/rpt_family="Alu"
13657..13857
/rpt_family="B2"
14075..14147
/rpt_family="ID"
14201..14226
/rpt_family="B4"
14227..14357
/rpt_family="Alu"
14479..14860
/rpt_family="L1"
17823..17954
/rpt_family="Alu"
18495..18597
/rpt_family="Alu"
18545..18631
/rpt_family="BCL_MM"
18973..19078
/rpt_family="B4"
19119..19242
/rpt_family="Alu"
19243..19352
/rpt_family="B4"
19432..19499
/rpt_family="ID"
22725..22836
/rpt_family="B4"
23994..24053
/rpt_family="B4"
24034..24194
/rpt_family="B2"
24230..24487
/rpt_family="L1"
26457..26650
/rpt_family="B2"
28112..28276
/rpt_family="MERR1_type"
29623..29699
/rpt_family="L2"
29973..30306
/rpt_family="ERVK"
30609..30705
/rpt_family="Alu"
31040..31148
/rpt_family="B4"
31927..31987
/rpt_family="Alu"
32001..32082
/rpt_family="L1"
32158..32257
/rpt_family="L1"
32837..32963
/rpt_family="Alu"
33965..34087
/rpt_family="B2"

```

```

repeat_region 34118. .34211
/rpt_family="Alu"
repeat_region 34226. .34383
/rpt_family="B4"
repeat_region 34746. .34857
/rpt_family="Alu"
repeat_region 35012. .35208
/rpt_family="B2"
repeat_region 36296. .36543
/rpt_family="MIR"
repeat_region 37116. .37236
/rpt_family="B4"
repeat_region 37323. .37461
/rpt_family="B4"
repeat_region 37722. .37853
/rpt_family="B4"
repeat_region 38215. .38354
/rpt_family="Alu"
repeat_region 39131. .39200
/rpt_family="Alu"
repeat_region 39392. .39414
/rpt_family="B2"
repeat_region 39415. .39496
/rpt_family="B4"

Query Match 2.9%; Score 39.4; DB 10; Length 186858;
Best Local Similarity 54.5%; Pred. No. 3.8;
Matches 79; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 1215 ACGTAGATACATAGAGTCATAGAAACATAAAAGGAGCTTGAAGAACCAATTCATA 1274
Db 130261 ATGTACAAACATATACTACACACAAATAAAANGTAAATTTGAAAAAGAGTAAATATCAT 130202

Qy 1275 GGGTCTCTCTTCTTTCGTCATCATCAAGATCATACATCAAAACGAGGAAGCTTTCT 1334
Db 130201 TAAGTCAAAACACACAGGTACACACACACATTAACACAAACAAACAAACAA 130142

Qy 1335 ATCTTCCTATAGCAATTCCTCAAAA 1359
Db 130141 ACCCTCCCTAAACCAACCAAAAAA 130117

```

Search completed: June 10, 2004, 11:49:37
 Job time : 3799.21 secs

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 09:32:10 ; Search time 2139.6 Seconds

(without alignments)
15199.092 Million cell updates/sec

Title: US-10-070-386-2

Perfect score: 1089

Sequence: 1 taaactccatcatatagcgg.....actgcaatgagtcgcccggg 1089

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gsl1:*

29: gb_gsl2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	
1	42.8	3.9	1101	29	CNS0000D1	
C	2	42.2	3.9	1201	13	AX376097
C	3	41	3.8	440	28	AQ634841
C	4	40.2	3.7	1101	29	CNS0039G
						AL065414
						AX376097
						AQ634841
						AL063931
						Drosophila

C	5	40	3.7	1101	29	CNS014QD
	6	39.8	3.7	638	28	AQ385240
	7	39	3.6	807	28	BZ718506
	8	39	3.6	833	28	BZ417799
	9	39	3.6	894	29	CG107311
	10	39	3.6	973	29	CG038009
	11	38.8	3.6	505	14	CB943651
	12	38.4	3.5	492	29	CE323888
	13	38.2	3.5	823	13	BU374247
	14	38.2	3.5	839	29	CG865000
	15	37.8	3.5	586	10	AW384856
	16	37.8	3.5	595	13	BW240301
	17	37.6	3.5	1201	13	BX337130
	18	37.4	3.4	555	28	AZ943190
	19	37.4	3.4	660	9	A1488706
	20	37.4	3.4	743	9	A1486681
	21	37.4	3.4	842	29	BCA576716
	22	37.2	3.4	581	29	CB457309
	23	37.2	3.4	760	29	CG120814
	24	37.2	3.4	797	29	CNS00358
	25	37	3.4	458	29	EX002272
	26	37	3.4	552	13	BU497199
	27	37	3.4	624	29	CB610094
	28	37	3.4	1001	29	CNS01400
	29	36.8	3.4	524	14	CF180840
	30	36.8	3.4	560	28	AQ237122
	31	36.8	3.4	657	28	CC095789
	32	36.8	3.4	762	29	AG030225
	33	36.6	3.4	353	10	BF490341
	34	36.6	3.4	355	10	BE978369
	35	36.6	3.4	368	28	BZ761149
	36	36.6	3.4	413	14	CB207284
	37	36.6	3.4	587	10	AW441466
	38	36.4	3.3	366	10	AW765028
	39	36.4	3.3	411	12	BG553280
	40	36.4	3.3	442	10	BG023598
	41	36.4	3.3	481	12	BG413804
	42	36.4	3.3	490	12	BG348032
	43	36.4	3.3	492	13	BQ737566
	44	36.4	3.3	531	12	BG579410
	45	36.4	3.3	612	14	CF287145

ALIGNMENTS

RESULT 1	CNS0000D1	1101 bp	DNA	linear	GSS 03-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence TET3 end of BAC #				
DEFINITION	BACK01J16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL065414				
VERSION	AL065414.1	GI:4938827			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster (fruit fly)				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseogawa and Aaron Mammosier in Pieter de Jong's laboratory in the Department of				

REFERENCE 1 (bases 1 to 440)
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
 TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
 JOURNAL Map Building (1997)
 COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbeetigr.org
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search.html.
 Seq primer: SP6
 Class: BAC ends.

FEATURES

source Location/Qualifiers
 1..440
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="GBA:7683074"
 /db_xref="taxon:9606"
 /clone="RPCI-11-477M3"
 /sex="Male"
 /cell_type="Lymphocytes"
 /clone_lib="RPCI-11"
 /notes="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
 RPCI11 Human Male BAC Library"

ORIGIN

Query Match 3.8%; Score 41; DB 28; Length 440;
 Best Local Similarity 50.8%; Pred. No. 0.46;
 Matches 95; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
 QY 640 CACTGCTATGCTGGGTATGTAATCTCGCTAGTTTCATAAGGGGCGACATAGAAAGAATA 699
 DB |||||
 QY 401 CATTTGTGTCTAAGCCACNAGTCGTGTGTTATTTGTTATGGAGCTTAGAAGACTA 342
 DB |||||
 QY 700 CTTCACTGTCAGCATACGTAAGCTCTGTACATTTCTCACTGCAAAATTTCTGAAACAATGGAG 759
 DB |||||
 QY 341 ACACAACTACAGAAAAGGTAGTCTGTATGCGTTACAGAACATCACTATAAAATAGAAC 282
 DB |||||
 QY 760 AGCATTATGAATTAATGGAATCTCTCATATATAGTGGAAGAAACAGAGCGCCCTTTTA 819
 DB |||||
 QY 281 AACTGCTANAAGATCTTTTGGTCTTTTATCAAAAGTGAGCAACAATGAAGACTTATA 222
 DB |||||
 QY 820 TTATGAA 826
 DB |||||
 QY 221 TTATCAA 215

RESULT 4
 CNS0039G/c 1101 bp DNA linear GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
 DEFINITION BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL063921 GI:4941778
 VERSION AL063921.1
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Sukayota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)
 AUTHORS Genoscope.

TITLE

JOURNAL Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source Location/Qualifiers
 1..1101
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR08K10"
 /clone_lib="RPCI-98"
 /note="end : TET3"

ORIGIN

Query Match 3.7%; Score 40.2; DB 29; Length 1101;
 Best Local Similarity 14.7%; Pred. No. 1.4;
 Matches 60; Conservative 184; Mismatches 163; Indels 0; Gaps 0;
 QY 327 TGAATGGGTATTCATATATATGAGCTTGACCCGCCCAATATGTAAACAGACCAACAATG 386
 DB |||||
 QY 940 WDDWDWDKWDGWDGAKDKRADDGDDGDKDDGKDDDDTGTGDKDDDKDDWDKA 881
 DB |||||
 QY 387 GGTCACTGGCCCCCAGACATGTGGCTATATPAAGTACCTCTGTAGCAATCAGACTTAC 446
 DB |||||
 QY 880 KGTWGDATWAAATDWWGWDADWWTDAADDDWADDDWDADWAWKWDADAWAGARTA 821
 DB |||||
 QY 447 TGATAGACGCTCCCTATATATGCTATAAATAAGTCACCTACTAGAACTACCGACAGTGTG 506
 DB |||||
 QY 820 DRDWDGAGKGGGARKRRDRKRDADDDADDDADATTTTTRDIDDDWKKTD 761
 DB |||||
 QY 507 AAATCCGACAGTGTCTGTGTTGAACATGTCATGCTATATGAATGAATGAAGAAG 566
 DB |||||
 QY 760 WTRWAADRTWRDDDDDDRDAGTAGCKWRRTWKRRKRRDTRWDDADADDTARDRRRRG 701
 DB |||||
 QY 567 GTGTACGGGTAGTACGAATCTGTATGATCAATGAGTACGCTAGTGTGTAACAGCG 626
 DB |||||
 QY 700 DDGADAGKGTGKRRRRDRATWDRDADAWWTTTDTDDDDKRRRRKRRR 641
 DB |||||
 QY 627 GATCGGATCTAGCACTGCTATGCTGGGTATGTAATCTCGCTATGTTCTAAGCGGA 686
 DB |||||
 QY 640 RETTARAADWWTKAWDKWDMKTRADRDWMAADTWTDAKADRDWAKARWRARD 581
 DB |||||
 QY 687 CATGAAGAATACCTCAGTGTACGATACGTAAGCTCTGTACATT 733
 DB |||||
 QY 580 RARAARDRRTTKGTTTATTTTWAARAAMWMAWATTTATTTT 534

RESULT 5

CNS014QD/c 1101 bp DNA linear GSS 26-JUL-1999
 LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
 DEFINITION BACN12C15 of drosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL104479
 VERSION AL104479.1
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster


```

Seq primer: TP
Class: sheared ends.
Location/Qualifiers
  1. .807
    /organism="Zea mays"
    /mol_type="genomic DNA"
    /strain="B73"
    /db_xref="taxon:4577"
    /clone="ZMMBTal36L20"
    /notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
    CoT selected genomic DNA library"

ORIGIN
Query Match          3.6%; Score 39; DB 28; Length 807;
Best Local Similarity 48.8%; Pred. No. 2.7;
Matches 105; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 579 AGTACGAATCTGTATGATATCAATGATGAGTAAACAGCGGATCGGATCTA 638
    |||||
Db 115 AGTACGAATGTTACATATATATTTAAAGTCTGTAATCAAGGGTACTTTGAA 174

QY 639 GCATCTGCTATGCTGGGTATGTAATCTCGCTATGTTTCATPAGGGGACATAGAAAGAT 698
    |||||
Db 175 GCACCAATATTTTGGACCAATTCAAACATCATTTAAATTTAGACAAACAGAGGTAT 234

QY 699 ACCTCAGTGTGACATACGTAGCTGTACATTTCACTGCAATTTCTGAACAATGGA 758
    |||||
Db 235 AACTCACTTGCAGTGTGCAAGTCTCTTTTAAAGCTCATCAGCACCTGAACATTAGA 294

QY 759 GAGCATTATGAATACTAAATGGAATCTCTCATTA 793
    |||||
Db 295 AAGAGGTTCACAGTCAGCAATACAGACTTCAGTA 329

RESULT 8
BZ417799
LOCUS          833 bp      DNA      linear      GSS 10-DEC-2002
DEFINITION    if81f07.g2 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
VERSION       if81f07.5, genomic survey sequence.
ACCESSION     BZ417799
KEYWORDS      GSS.
SOURCE        BZ417799.1 GI:26361596
ORGANISM      Zea mays
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
              clade; Panicoideae; Andropogoneae; Zea.
              1 (bases 1 to 833)
              Rabinowicz,P.D., O'Shaughnessy,A.L., Ballja,V., Dedhia,N.,
              Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,
              Zutavern,T., McCombie,W.R. and Martienssen,R.A.
              Genomic shotgun sequences from Zea mays (methyl-filtered)
              Unpublished (2002)
              Contact: W. Richard McCombie
              Lita Annenberg Hazen Genome Sequencing Center
              Cold Spring Harbor Laboratory
              PO Box 100, Cold Spring Harbor, NY 11724, USA
              Tel: 516 367 8884
              Fax: 516 367 8874
              Email: mcombie@cshl.org
              Plate: if81 row: f column: 07
              Seq primer: -21M13UnivRev
              Class: shotgun
              High quality sequence stop: 833.
              Location/Qualifiers
                1. .833
                  /organism="Zea mays"
                  /mol_type="genomic DNA"
                  /cultivar="B73"
                  /db_xref="taxon:4577"
                  /clone="if81f07"
                  /lab_host="DH5a"

FEATURES
  source
    1. .833
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBTal36L20"
      /notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
      CoT selected genomic DNA library"

ORIGIN
Query Match          3.6%; Score 39; DB 29; Length 894;
Best Local Similarity 48.8%; Pred. No. 2.8;
Matches 105; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 579 AGTACGAATCTGTATGATATCAATGATGAGTAAACAGCGGATCGGATCTA 638
    |||||
Db 604 AGTACGAATGTTACATATATATTTAAAGTCTGTAATCAAGGGTACTTTGAA 663

QY 639 GCATCTGCTATGCTGGGTATGTAATCTCGCTATGTTTCATPAGGGGACATAGAAAGAT 698
    |||||
Db 664 GCACCAATATTTTGGACCAATTCAAACATCATTTAAATTTAGACAAACAGAGGTAT 723

QY 699 ACCTCAGTGTGACATACGTAGCTGTACATTTCACTGCAATTTCTGAACAATGGA 758
    |||||
Db 724 AACTCACTTGCAGTGTGCAAGTCTCTTTTAAAGCTCATCAGCACCTGAACATTAGA 783

QY 759 GAGCATTATGAATACTAAATGGAATCTCTCATTA 793
    |||||
Db 784 AAGAGGTTCACAGTCAGCAATACAGACTTCAGTA 818

RESULT 9
CG107311/c
LOCUS          894 bp      DNA      linear      GSS 20-AUG-2003
DEFINITION    PUJEA72TB ZM 0.6-1.0 KB Zea mays genomic clone ZMMETA0647K23,
              Genomic survey sequence.
ACCESSION     CG107311
VERSION       CG107311.1 GI:33990189
KEYWORDS      GSS.
SOURCE        Zea mays
              Zea mays
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
              clade; Panicoideae; Andropogoneae; Zea.
              1 (bases 1 to 894)
              Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
              Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
              Bennetzen,J.
              Maize Genomics Consortium
              Unpublished (2003)
              Other GSSs: PUJEA72TD
              Contact: Cathy Whitelaw
              TIGR
              9712 Medical Center Drive, Rockville, MD 20850, USA
              Tel: 301-838-5843
              Fax: 301-838-0208
              Email: whitelaw@tigr.org
              Seq primer: TR
              Class: sheared ends.
              Location/Qualifiers
                1. .894
                  /organism="Zea mays"
                  /mol_type="genomic DNA"
                  /strain="B73"
                  /db_xref="taxon:4577"
                  /clone="ZMMETA0647K23"
                  /notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
                  CoT selected genomic DNA library"

FEATURES
  source
    1. .894
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMETA0647K23"
      /notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
      CoT selected genomic DNA library"

```

Best Local Similarity 48.8%; Pred. No. 2.9;
Matches 105; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 579 AGTACGAATCTGTATGATATCAATAGGTAGCAGTGTGTAACACGCGGATCGGATCTA 638
Db 225 AGTAACATATGTGTATCATATATATTTTAAAGTCTGTAAATCAAGGGTACTTTTGA 166
QY 639 GCACTGCTATGTCTGGTATGTAATCTGCTATGTCATAGGGCGCATAGAAAGAT 698
Db 165 GCACCATATTTTGGAACTAATCAACATCAATTAATTTAGACAAACAGACGTAT 106
QY 699 ACTGAGTGCACATAGTACGTCTGTACATTTTCACTGCGAAATTTCTGAACAATTGA 758
Db 105 AACTCACTTGCAGTGTCAATGTCTCTTTTAAAGTCTCATCAGCGACCTGAACATTAAGA 46
QY 759 GACCATTTGAAATCAATCAATGCACTCTCTCAATTA 793
Db 45 AAGAGTFCAGTCAACCAATACAGACTTCAGTA 11

RESULT 10
CG038009
LOCUS
DEFINITION PUF0241TD ZM 0.6 1.0 KB Zea mays genomic clone ZM5BTA0713G10,
genomic survey sequence.
ACCESSION CG038009
VERSION CG038009.1 GI:33910165
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 973)
AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other GSSs: PUF0241TB
Contact: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.

FEATURES
source
1..973
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZM5BTA0713G10"
/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
CoT selected genomic DNA library"

ORIGIN
Query Match 3.6%; Score 39; DB 29; Length 973;
Best Local Similarity 48.8%; Pred. No. 3.1; Indels 0; Gaps 0;
Matches 105; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 579 AGTACGAATCTGTATGATATCAATAGGTAGCAGTGTGTAACACGCGGATCGGATCTA 638
Db 397 AGTAACAATATGTTACATATATATTTTAAAGTCTGTAAATCAAGGGTACTTTTGA 456
QY 639 GCACTGCTATGTCTGGTATGTAATCTGCTATGTCATAGGGCGCATAGAAAGAT 698
Db 457 GCACCATATTTTGGAACTAATCAACATCAATTAATTTAGACAAACAGACGTAT 516
QY 699 ACTGAGTGCACATAGTACGTCTGTACATTTTCACTGCGAAATTTCTGAACAATTGA 758

Db 517 AACTCAGTTCAGAGTTCGAATGCTCTTTTAAAGTCTCATCAGCGACCTGAACATTAAGA 576
QY 759 GAGCATTTATGAATCTAAATGAAGAACTCTCTCAATTA 793
Db 577 AAGAGGTCCAGTTCAGCAAAATACAGACTTCAGTA 611

RESULT 11
CB943651
LOCUS
DEFINITION CB943651 505 bp mRNA linear EST 29-APR-2003
IMAGE:6933934 5', mRNA sequence.
ACCESSION CB943651
VERSION CB943651.1 GI:30191094
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 505)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov
Tissue Procurement: Drs. Donald Brown and Liqian Cai
cDNA Library Preparation: CLONTECH
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM3249 row: 0 column: 21
High quality sequence stop: 459.

FEATURES
source
1..505
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clones="IMAGE:6933934"
/dev_stage="metamorphosis stage 62"
/clone_lib="NICHD XGC Tad2"
/note="Organ: Developing Tadpole; Vector: pDNR-LIB;
Site 1: SfiI; Site 2: SfiI, 5' and 3' adaptors were used
in cloning as follows: 5' adaptor sequence:
5'-ATTCTAGAGCGCGCGCGCGCATG-df(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.7 kb
(range 0.8-3.0 kb). 15/15 colonies contained inserts by
PCR. This library was enriched for full-length clones and
was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN
Query Match 3.6%; Score 38.8; DB 14; Length 505;
Best Local Similarity 55.2%; Pred. No. 2.4;
Matches 73; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 683 GCGACATAGAAGATACCTCAGTGTGTCAGCATAGTTCAGTCTCTCACTTTCAGTCAAA 742
Db 351 GCACATATACAGATGTCACCTCAAGTCAAGAAAGAAACAGACCTGCGACTGTACG 410
QY 743 TTTCTGAACAATTGGAGAGCATTTATGAATACTAAATGGAACCTCTCAATTATAGTGA 802
Db 411 TTTGTTTACAGCGCATCATATTTATTACTTTCTTAATAATAAACACATCGCGTTAAGTCAA 470
QY 803 AACAGAGCGC 812
Db 471 AAAAGACCAC 480

RESULT 12
CE323888
LOCUS
DEFINITION
tigr-gss-dog-17000360301767 Dog Library Canis familiaris genomic,
Genomic survey sequence.
ACCESSION
CE323888.1 GI:36133243
VERSION
CE323888
KEYWORDS
GSS.
SOURCE
Canis familiaris (dog)
ORGANISM
Canis familiaris
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Bukaryota; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE
1 (bases 1 to 492)
AUTHORS
Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,
Rusch, D.B., Deicher, A.L., Pop, M., Wang, W., Fraser, C.M. and
Venter, J.C.
TITLE
The dog genome: survey sequencing and comparative analysis
JOURNAL
Science 301 (5641), 1898-1903 (2003)
MEDLINE
22875432
PUBMED
14512627
COMMENT
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers
1..492
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match 3.5%; Score 38.4; DB 29; Length 492;
Best Local Similarity 47.5%; Pred. No. 3.2;
Matches 114; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
QY 669 CTATGTCATAGGGGACATAGAAAGATACCTCAGTGTGACATACGTAAGCTCTGA 728
Db 130 CTGCTTTTATATGGGAATAGACAAGTATCTCTGTGTCATTATCCCTCTCCACATT 189
QY 729 CATTCTACTGCAATTTCTGCAAAATGGAGCATATGAAATCTAAATGGAATCTCT 788
Db 190 GAATCAGCGTCCCTTCCCAATACAGTAAATAAGAAAAGAAATACATATGTAAACCA 249
QY 789 CATTATAAGTGGAAAAACAGAGCGCCCTTTTATTATGAAACAGAGCGGTCAAGAACGTCTT 848
Db 250 CAAAAGAGAAAAGAAAGAAATGTCATTTTACAGAAATATATATTTGTGAGCAAGCAGA 309
QY 849 TCACTCATCAGAGCGGTTCCTCCATCAGATCATCTTCCTTGAACCATGTTCTGCGAT 908
Db 310 AAAGCAAAATATATGAATTTAAACAGAGCGCCCTTGGGTGGCTCAGTGGTGTGACAT 369

RESULT 13
BU374247/c
LOCUS
DEFINITION
603588074f1 CSRQCHN74 Gallus gallus cDNA clone CHEST547c20 5', mRNA
sequence.
ACCESSION
BU374247
VERSION
BU374247.1 GI:259882248
KEYWORDS
EST.
SOURCE
Gallus gallus (chicken)
ORGANISM
Gallus gallus
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

REFERENCE
1 (bases 1 to 823)
AUTHORS
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
TITLE
A Comprehensive Collection of Chicken cDNAs
JOURNAL
Curt. Biol. 12 (22), 1965-1969 (2002)
MEDLINE
22335534
PUBMED
12445392
COMMENT
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1..823
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="CHST547c20"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSRQCHN74"
/note="Organ: kidney + adrenal; Vector: pBluescript II
KS(+); Site 1: EcoRI; Site 2: NotI; This normalized
library was constructed from 1 million independent clones.
cDNA synthesis was initiated using an oligo(dT) primer,
using methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

ORIGIN
Query Match 3.5%; Score 38.2; DB 13; Length 823;
Best Local Similarity 51.5%; Pred. No. 4.9;
Matches 88; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 562 AGAAGGTGACGGGTTAGTACGAATCTGTATGATTAATCAATGTAGCAGTGTGTTAA 621
Db 760 ACATGGCTCTTGTGTTTATAGTCTATCAGAGATAGGATGGGGCAAAAAGCAAG 701
QY 622 CAGCGATCGGATCTAGCATCTGTCTGGGTATGTATCTCTGGCTATCTTCTATAG 681
Db 700 CAGCAAAAGATGCACAGGAATCTGAACATTTGCCAAGAAATTCATGTTCTTCAGAAC 641
QY 682 GCGACATAGAAAGATACCTCAGTGTGACGATAGTAAAGCTCTCTACATT 732
Db 640 TGCATCTAGAAATTAACCTCAGTCTCTGATACGCAATCTGTGATT 590

RESULT 14
CG865000/c
LOCUS
DEFINITION
ZMMBBc0273K14f ZMMBc (Scorl) Zea mays subsp. mays genomic clone
ZMMBBc0273K14 5', genomic survey sequence.
ACCESSION
CG865000
VERSION
CG865000.1 GI:38441215
KEYWORDS
GSS.
SOURCE
Zea mays subsp. mays (maize)
ORGANISM
Zea mays subsp. mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1 (bases 1 to 839)

AUTHORS Bharti, A.K., Young, S., Kavchok, S., Keizer, G., Bronzino, A.C.,
 Zohovetz, V., Fuks, G., Yu, Y., Wang, R. and Messing, J.
TITLE Sequencing of the maize genome at PGIR (2003c)
JOURNAL Unpublished (2003)
COMMENT Contact: Bharti, A.K.
 Dr. Joachim Messing's lab

The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
 University
 190 Frelinghuysen Road, Piscataway, NJ 08854, USA
 Tel: 732 445 3801
 Fax: 732 445 5735
 Email: bharti@waksman.rutgers.edu
 Seq primer: T7
 Class: BAC ends
 High quality sequence start: 507.

FEATURES

source
 1. .839
 Location/Qualifiers
 /organism="Zea mays subsp. mays"
 /mol_type="genomic DNA"
 /cultivar="B73"
 /sub_species="mays"
 /db_xref="taxon:4578"
 /clone="ZMBBc0273K14"
 /lab_host="E. coli DH10B"
 /clone_lib="ZMBBc (EcoRI)"
 /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI"

ORIGIN

Query Match 3.5%; Score 38.2; DB 29; Length 839;
 Best Local Similarity 50.3%; Pred. No. 5;
 Matches 94; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
 QY 632 GCATCTAGCACTGCTATGCTGGGTATGTAATCCTGGCTATGTTCTAAGGGGCACATAG 691
 DB 815 GCATCATGTGGGTTTACTTGCTGCGCATTTGGTATATAAATAATGAGATATAA 756
 QY 692 AAGAATACCTCAGTGTACGATAGCTCTGTACATTTTCATCGCAAAATTTCTGAAC 751
 DB 755 ACAATATATTAAATGACATCTAGAATTTAACTTCTACATTTGAAACCAAACTCTAAAG 696
 QY 752 AATTGAGAGACATTATGAATATCTAAATGGAATCTCTCATTAATAGTGAACACAGCG 811
 DB 695 AAAGAGAGAGGGAAGGAATCTTATATATAGTCCAGAGAAAGTTTATATAGATAT 636
 QY 812 CCTTTT 818
 DB 635 CCCTAT 629

RESULT 15

AW384856
 LOCUS RC3-HT0419-301299-011-d02 HT0419 Homo sapiens cDNA, mRNA sequence.
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 586)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

HCSP http://www.ludwig.org.br/ORESTES.
 The FAPESP/LiCR Human Cancer Genome Project
 Unpublished (1999)
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LiCR Human Cancer Genome

Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?c1=RC3&t2=RC3-HT0419-
 301299-011-d02&t3=1999-12-30&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 9
 High quality sequence stop: 585.

FEATURES

source
 1. .586
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="HT0419"
 /note="Organ: head neck; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

ORIGIN

Query Match 3.5%; Score 37.8; DB 10; Length 586;
 Best Local Similarity 54.7%; Pred. No. 5.4;
 Matches 75; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
 QY 513 GACAGTGTCTGCTGTTGAACATGCTCATGTCTATATGAATGAATAAGAAAGGTGCA 572
 DB 99 GACAGTGTCTGCTGCTATAGGAGGTGTACAATACATATTTGTTGATGAGAAAGATCA 158
 QY 573 CGGGTTAGTACGAATCTGTATGATATCAATGTTAGTACGATGTTGTTAAACACCGGATCG 632
 DB 159 TGGATTAAATGAAGAGCTTCATGCTCTACAAAGAACTGATACACAGTAAATCAGGAAGGG 218
 QY 633 GATCTAGCACTGCTATG 649
 DB 219 GATCAAGAACTCTCTATG 235

Search completed: June 10, 2004, 13:10:48
 Job time : 2144.6 secs

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 08:08:35 ; Search time 332.836 Seconds
(without alignments)
13899.608 Million cell updates/sec

Title: US-10-070-386-2
Perfect score: 1089
Sequence: 1 taaactccatcatatagcg...actgcgaatagtcgcccggg 1089

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseg 29Jan04:.*
1: Genesegm1980s:.*
2: Genesegm1990s:.*
3: Genesegm2000s:.*
4: Genesegm2001as:.*
5: Genesegm2001bs:.*
6: Genesegm2002s:.*
7: Genesegm2003as:.*
8: Genesegm2003bs:.*
9: Genesegm2003cs:.*
10: Genesegm2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1089	100.0	1089	5	Aaf79835 Mycelia s
2	47	4.3	2000	7	Ada71938 Rice gene
3	44.8	4.1	2000	7	Ada71938 Rice gene
4	44	4.0	14041	4	Aah48024 Internal
5	41.2	3.8	10732	3	Aaal0594 Gene enco
6	40.8	3.7	276	9	Add17664 DNA (Seqi
7	36.2	3.3	49998	2	Aax23518 Human kid
8	36.2	3.3	177380	7	Acf62751 Cancer ba
9	36.2	3.3	177380	7	Adb20870 MRP1 base
10	36.2	3.3	177380	9	Adb87959 Human UGT
11	36.2	3.3	177380	9	Adb87959 Human MDR
12	36.2	3.3	177380	9	Adb96942 Human MDR
13	35.4	3.3	7025	6	Adb92133 Human MDR
14	35.4	3.3	7025	6	Aak40059 Human che
15	35.2	3.2	2604	4	Aas63350 Chemical
16	35.2	3.2	2604	4	Aah16164 Human cdn
17	35.2	3.2	2604	4	Aht07177 Human ova
18	35.2	3.2	4020	4	Aal06054 Human rep
19	35	3.2	4020	4	Abl98619 Human tes
20	34.8	3.2	4072	4	Abl12544 Drosophil
21	34.4	3.2	40308	4	Aak68184 Human imm
22	34.4	3.2	1370	6	Abq69040 Listeria
23	34.4	3.2	5818	6	Abq71014 Listeria
24	34.4	3.2	110000	6	Abq67196_4

Continuation (5 of

C 24	34.4	3.2	110000	6	ABQ69245_27	Continuation (28 o
C 25	34.4	3.2	143306	6	ABK49586	Abk49586 Human tra
C 26	34.2	3.1	576	4	AAI16818	AAI16818 Probe #67
C 27	34.2	3.1	576	4	ABA60543	ABA60543 Human foe
C 28	34.2	3.1	576	4	AAI40432	AAI40432 Probe #91
C 29	34.2	3.1	576	4	ABA28703	ABA28703 Probe #71
C 30	34.2	3.1	576	4	AAK34711	AAK34711 Human bon
C 31	34.2	3.1	576	4	AAK08823	AAK08823 Human bra
C 32	34.2	3.1	576	4	ABS34484	ABS34484 Human liv
C 33	34.2	3.1	576	6	ABS09300	ABS09300 Human imm
C 34	34.2	3.1	576	4	AAK79610	AAK79610 Human imm
C 35	34.2	3.1	5835	4	AAK72786	AAK72786 Human imm
C 36	34.2	3.1	5835	4	AAK79611	AAK79611 Human imm
C 37	34	3.1	5449	5	AAK58587	AAK58587 DNA encod
C 38	33.8	3.1	366	7	ACA22237	ACA22237 Prokaryot
C 39	33.6	3.1	432	7	ABX53039	ABX53039 Bovine ES
C 40	33.4	3.1	2853	7	ACA23231	ACA23231 Prokaryot
C 41	33.4	3.1	110000	2	AAK20248_00	AAK20248 Borrelia
C 42	33.2	3.0	562	4	AAK37880	AAK37880 Human bon
C 43	33.2	3.0	1048	6	ABN74166	ABN74166 Bovine em
C 44	33	3.0	1957	6	ABS55349	ABS55349 cDNA enco
C 45	33	3.0	19736	6	ABA92158	ABA92158 Human pro

ALIGNMENTS

RESULT 1
AAF79835
ID AAF79835 standard; DNA; 1089 BP.
XX
AC AAF79835;
DT 30-MAY-2001 (first entry)
XX Mycelia sterilia terminator sequence.
DE Promoter; terminator; regulatory region; filamentous fungus;
XX protein production; ds.
XX Mycelia sterilia.
PN WO200118219-A1.
PD 15-MAR-2001.
XX
PF 07-SEP-2000; 2000WO-JP006104.
XX
PR 07-SEP-1999; 99JP-00252851.
XX (MEIJ) MEIJI SEIKA KAISHA LTD.
XX Watanabe M, Murakami T;
XX WPI; 2001-235202/24.
XX New promoter and terminator functioning synchronously, useful for
XX regulating expression of endogenous gene in filamentous fungi,
XX particularly transformed Mycelia sterilia capable of producing high
XX yields of target protein/substance.
XX Claim 6; Page 20-22; 26pp; Japanese.
XX The present invention provides the sequences of a promoter and terminator
XX from Mycelia sterilia which can be used to regulate the expression of an
XX endogenous gene in filamentous fungi, particularly transformed Mycelia
XX sterilia, to produce useful target proteins or other substances. The
XX present sequence is the terminator of the invention
XX
XX Sequence 1089 BP; 298 A; 248 C; 247 G; 296 T; 0 U; 0 Other;
Query Match 100.0%; Score 1089; DB 5; Length 1089;
Best Local Similarity 100.0%; Pred. No. 0;


```

Db 849 YTMVTCYAMTCAKCYKAMTKMTWACAWRATSWRWRAMGVRWVKYKOKRAYWNRW 790
Qy 482 CACTACTAGACTACGACAGTGTGAATCCGACAGTGTCTGCTGTGTGACATGTCAT 541
Db 789 RWCWAGARWKSRYRKKYATRYKWMAMTWMSWRWKSRYRWSGKGRWRSAN 730
Qy 542 GTCTATGATGAATAAGAGAGAGTGTGACGGGTAGTAGTACCAATCTGTATGATAATCA 601
Db 729 RYCSRMCAKTKYASSARWTKAKESYRWRWYWRKGTWYRYRWSRMTARMKR 670
Qy 602 ATGTAGAGTATGATGTAACA----GGGATCGGATCTAGCACTGTATGTCGGTA 657
Db 669 RHWAGASMKCMWYWRGARSWMYSKYSKACCKKTRTYNTSYMTGMYGYSYKSMW 610
Qy 658 TGTATCTCGCTATGTTTCATAAGCGGACATAGAAATACCTCAGTGTGACATACG 717
Db 609 TSXNSYKMTCTMTYSKMGSTRSKMGWSGMSBMYRWMKWEKRYMYMKWCTWR 550
Qy 718 TAAGCTGTGATCTTCACTGCAAAATTTCTGAACAAATTTGGAGAGCAATATGAATACTAA 777
Db 549 RCMCYRWGYTMYTTSRSMYTYGRYKARYTSKRYMYKRYKYCHWYTYGYMYKCSYMYR 490
Qy 778 ATGGAACCTCATTATAGTGAAGAGAGAGCGCCCTTTATTATGAAACAGAGAGTC 837
Db 489 YGYCKKCKCCYAMCKAAYSKMMYTRKYSKWRMSTKTNWSWYKCKRSMYKAGC 430
Qy 838 AAGAACGCTTTCAACGTCATCAGAGCGGTTCCATCCAGATCATACTTCCCTTGAACA 897
Db 429 YGCRMTYCSYGMKWTYMGSYKYSRCYKYMRTMYKGMWYMYYSAYSSNMTWYTYA 370
Qy 898 TGTCTCGCATTCAGAACTGATGAGTGAAGAACCGTCCAGGGTTCCTGTCTATCCCTTG 957
Db 359 KYMYKYKRRGTMSYKYSKYYCTWYKMKRCYRWRMOMRKKYKRCYCWRYAT 310
Qy 958 CGTCCCTTCGAATAAATCGTATTACCA-TTTTCTTTCGACGCGCGGTCAACGTGAGCG 1016
Db 309 CYWCCYKRWGYSRSMRTAGKWRMSWGRWCRSYKMYKMYKMKWKSYYMSYGVARS 250
Qy 1017 ACGTGCCACGTTGAGTCCACATGACACAGTGGATGTCATCCACGCC 1065
Db 249 SGTWRSAAKRTYKGYSTRSRRAKWRACRMYSACRRYSRTSYCGCSYC 201

```

RESULT 3

```

ID ADA71938 standard; DNA; 2000 BP.
AC ADA71938;
DT 20-NOV-2003 (first entry)
DE Rice gene, SEQ ID 5263.
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
OS Oryza sativa.
PN WO2003000898-A1.
PD 03-JAN-2003.
PF 22-JUN-2001; 2001WO-IB001105.
PR 22-JUN-2001; 2001WO-IB001105.
PA (SYGN) SYNGENTA PARTICIPATIONS AG.

```

Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,
 Katagiri F, Quan Y, Tao Y, Whitham S, Xie Z, Zhu T, Zou G,
 WPI; 2003-175290/17.

XX Identifying at least one gene involved in plant resistance or response to
 PT pathogenic infection for conferring resistance or tolerance to a plant to
 PT bacterial, fungal or viral infection by determining or detecting plant
 XX gene expression.

Claim 27; SEQ ID NO 5263; 899pp; English.

XX The present invention relates to a method (M1) for identifying genes
 XX involved in plant resistance or response to pathogenic infection. M1
 CC comprises identifying a gene whose expression is significantly altered in
 CC the incompatible interaction of plant gene expression relative to
 CC the expression of the gene in an uninfected plant, in a mutant plant that
 CC does not express a gene associated with response to pathogenic infection,
 CC or in a corresponding incompatible or compatible interaction. (M1) is
 CC useful for conferring resistance to resistance or tolerance to a plant to
 CC bacterial, fungal or viral infection. The present sequence was used to
 CC illustrate the invention.

XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 4.1%; Score 44.8; DB 7; Length 2000;

Best Local Similarity 11.3%; Pred. No. 0.0024;

Matches 104; Conservative 398; Mismatches 410; Indels 12; Gaps 4;

```

Qy 18 CGTTGGCTGAAGAGGAGAGAGAGAGATCAGTCTTTTCCAGCAGCTTGAGATCT 77
Db 73 SGRGTGWSGSGYSGKMKRYKRSKWRGRGRGRMRMRMRMRMRMRMRMRMRMRMR 132
Qy 78 GATCTGTGTTTCACTCTCAGTAATCTGTGTGGAAGTTACACTTCTGCTCTCCCTTAC 137
Db 133 GRWVGSRMSYMWYVARGCGSKKSKSGSWGKTCRGRGSGWSGWSGAKYSKMS 192
Qy 138 CAGCCTTCAGGCAACCAAGATGTTAGAGTTCGCTCATTTATATAGCTCTGGCGA 197
Db 193 KRWMSGCRSGGCRSAYSRVYGTSRKYTKYKMYTYSASRCWRATWITSWACSYT 252
Qy 198 TGACTAGCATTTAGGCGATGACGACATGCTGCTACTGCTGCTGCTGTTGGTTAGTTA 257
Db 253 WCSRKSRMMWKRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMR 312
Qy 258 CCTAGCTAGCAATATCAACATACAAATGTGGTTTCCACATGTCTAGCTGGTTCTACC 317
Db 313 YWGRGYMSRMWYKMYVRYGKMKRWGWRMRMRMRMRMRMRMRMRMRMRMRMRMR 372
Qy 318 TAGTCTGAGTGAATGGTAAATTTGATATATGAGCTTGACCCCGCAATATTGTAACAGAG 377
Db 373 RWAQKSRSTRSRKRRKWKRRKRYKMRGYSRMRMRMRMRMRMRMRMRMRMRMRMR 432
Qy 378 CCAACAATGGGTCACTCGGCCCGCCAGACATGTGGGTATATAAGCTACCTGTCTA----- 432
Db 433 TCRMKSYGMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMR 492
Qy 433 --GCAATCAGACTTACTAGTAGAAACGTCGCCCTATATGTCATATAAATAAGTACTACTAG 490
Db 493 RSGMKRCKRRRWRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMRMR 552
Qy 491 AACTAGCAGATGTGAATAATCCGACAGTGTCTGCTGTGGAACATGTCTATGTCTATATG 550
Db 553 GMMKRYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYK 612
Qy 551 AATGAATAAGAGAGAGTGTGACGGGTAGTAGCAATCTGTATGATATAATCATGTTAGCA 610
Db 613 SMRSRKRCKRCKRCKRCKRCKRCKRCKRCKRCKRCKRCKRCKRCKRCKRCKRCKR 671
Qy 611 GTGATGTAAACAGCGGATCGGATCTAGCACCTGCTATGCTGGGTATGTAATCCCTGGCT 670
Db 672 SKTYAKYSYWRYYRANCMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMY 731
Qy 671 ATGTTCAAGGCGCATAGAAAGATACCTCAGTGTGACGATACGATAGCTCTGTA-C 729
Db 732 SWYKCKGKYSRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 791

```


WFI; 2000-342371/30.
 P-PSDB; AAY85179.
 A gene encoding a cellulose synthetic equipment - for the improvement in the amount of cellulose synthesized in a plant body.
 Claim 2; Page 14-21; 32pp; Japanese.
 This sequence represents a gene encoding a subunit of the cellulose synthase complex of *Vigna angularis*. The invention relates to subunits of cellulose synthetic equipment, that can be used to increase the amount of cellulose synthesized by a plant. The proteins and genes encoding them can also be used to improve the properties of the cellulose being produced by a plant
 Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 0 U; 2251 Other;
 Query Match 3.8%; Score 41.2; DB 3; Length 10732;
 Best Local Similarity 15.0%; P-red. No. 0.089;
 Matches 65; Conservative 167; Mismatches 200; Indels 0; Gaps 0;
 QY 640 CACTGCTATGTCGGGTATGTAATCTCGCTATGTTCAATAGGGCGACATAGAAGAATA 699
 DB 10187 SNTGCTTYSDDGTCVCSRCRTBSYDAGNSPCRCSTYSCYTDACSTYASTDNCSSRSTTS 10128
 QY 700 CCFCACTGTCAGATACGTAAGCTCTGTATCATTTCACTGCAAAATTTCTGAACAATTCGAG 759
 DB 10127 TCTYAKYSTSTBNCTCYSRSGNSATKAYSTTTTAYDARCTDNTSTRCYASTYDAS 10068
 QY 760 AGCATATGAGAAATACATAATGGAATCTTCATATATAAGTGGAAAAACAGAGCGCCCTTTTA 819
 DB 10067 RRCYSTYASRGTSDDCCTBSRYSCYSDASRYANCDTBCYTTBAKYARCYDAYAKR 10008
 QY 820 TTATGAACAGAGAGCGTCAAGAACGCTTTTCAACGTCATCGAGCGGTTCCATCCAGATC 879
 DB 10007 CNSTSRAYRSDTCTYCRCRNCTYSYSSTYRATTTBTYCTCTBCRRYSRY 9948
 QY 880 ATACTTTCCCTTGAACCATGTTCTCGCATTCAGATCGTAGGATGGAACCGTCCAGG 939
 DB 9947 SSTCNCYSYCCYTSRYSSSTTNTSKCTTYSNSTTBTBYSYSSTYSRGSYSSRGS 9888
 QY 940 TTGCGTGTCATTCCTCGCTTGGCTTGCATGAATAAATCGTATTACCATTTCTTTCGCAGC 999
 DB 9887 NCYINSTNCYDASTSDTBYSRCCTYSYSSTSDTSDTSTNTTBSDCYTTTTBSRSTSD 9828
 QY 1000 GCGGGTCAACGTGAGCGAGCTGCCACGTTGGAGTCCCAATGACGAGTCATGTCATC 1059
 DB 9827 SNTYRCSRDYDABSDNSTWCYDASRTBTBSTNCYARCYTBDACSRDSTYSRSGY 9768
 QY 1060 CAGGCACACTCAC 1071
 DB 9767 DANSTSRYSRY 9756
 RESULT 6
 ADD17664/c
 ID ADD17664 standard; DNA; 276 BP.
 XX
 AC ADD17664;
 AC
 XX
 DF 15-JAN-2004 (first entry)
 DE DNA (SeqID 1732) that confers an altered visual phenotype in plants.
 XX
 KW ds; visual phenotype; plant; architecture; leaf surface; chlorotic;
 KW bleaching; etching; wet leaf; stunting; elongation; texture;
 KW agronomic trait; growth regulation; dwarf variety; insect resistance;
 XX heat stress; transgenic.
 XX
 OS Unidentified.
 XX
 PN WO2003020741-A1.
 XX

```

XX OS Homo sapiens.
XX KW W09911799-A2.
XX PD 11-MAR-1999.
XX PF 02-SEP-1998; 98WO-US018426.
XX PR 02-SEP-1997; 97US-0057854P.
XX PA (MEDI-) MEDICAL COLLEGE GEORGIA RES INST.
XX PI Ryan JW, Sprinkle TJC, Venema RC;
XX DR WPI; 1999-205193/17.
XX PT Nucleic acid encoding human aminopeptidase P.
XX PS Claim 13; Page 109-139; 201pp; English.
XX CC This invention describes the isolation of a novel human aminopeptidase P
CC (Amp). This protein is used to produce recombinant Amp and can be used
CC for gene therapy for treating Amp-deficiency conditions. Its fragments
CC are used as primers and probes to identify patients with homozygous and
CC heterozygous Amp deficiency, including prenatal diagnosis (patients
CC defective in Amp are at risk of developing angioedema if treated with
CC angiotensin-converting enzyme inhibitors), also as antisense inhibitors
CC in cases of excessive Amp expression. The product of the invention is
CC also used to identify Amp-expressing sequences in other animals and to
CC generate transgenic animals, and comparisons of genomic sequences are
CC used to detect mutations. Amp inhibitors are potentially useful as
CC antihypertensive agents and to prevent or treat arterial (re)stenosis or
CC atherosclerosis. The structure of Amp is used to design synthetic
CC substrates, e.g. for use in Amp assays. Amp, which hydrolyzes N-terminal
CC imido bonds, can be used to degrade industrial protein feeds to free
CC amino acids, to degrade proteinaceous wastes, as additives in enzyme
CC formulations used to treat malabsorption syndrome and for studying its
CC biological role. Antibodies against Amp are used in immunochemical
CC methods to study Amp distribution
XX SQ Sequence 49998 BP; 12605 A; 11725 C; 11351 G; 14317 T; 0 U; 0 Other;

Query Match 3.3%; Score 36.2; DB 2; Length 49998;
Best Local Similarity 48.8%; Pred. No. 8.8;
Matches 98; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 380 AACATGGCTCACCCTGGCCCCCAGACATGTGGCTATATAGCTACCTCTCTAGCAATCA 439
DB 23385 AAAATGGACTAAATATTTGAACAGACAAATCCACAAATAAGATATCTGGATGGCAATA 23326

QY 440 GACTTACTGATAGAACGTCCTCCCTATATGTCTATAAATTAAGTCACCTACTAGAACTACCGA 499
DB 23325 AACACATATAAGATGCTTTCCTCAATTAGCCATTAGGAGATATAAAATTAACACAGT 23266

QY 500 CAGTGTGAATCCGACAGTGTCTGGTCTGTGTAACATGTCATGTCTATATGATGATATAA 559
DB 23265 GAGATACCACTACAGATCTATTTCGAATGGCTAAAGTTAAAAAGACTGACTGAAGTACTAT 23206

QY 560 GAAGAAGGTGTACGGGTAG 580
DB 23205 TCCAAATTATGTGAGGTAGTAG 23185

RESULT 8
ACF62751/c
ID ACF62751 standard; DNA; 177380 BP.
XX AC ACF62751;
XX AC ACF62751;
XX 08-OCT-2003 (first entry)
DT Cancer based on CYP3A5 related polynucleotide SEQ ID NO:683.
DE

```

```

XX CC Cancer; CYP3A5; irinotecan; pharmaceutical; malignant glioma;
XX KW cytochrome p450; subfamily IIIA; nifedipine oxidase; polypeptide 5;
XX KW cytosolic; gene; ds.
XX OS Unidentified.
XX XX W02003013534-A2.
XX PD 20-FEB-2003.
XX PF 23-JUL-2002; 2002WO-EP008219.
XX PR 23-JUL-2001; 2001EP-00117608.
XX PR 24-MAY-2002; 2002EP-00011710.
XX PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX PI Heinrich G, Kerb R;
XX WPI; 2003-268144/26.
XX CC New use of irinotecan for preparation of compositions for treating cancer
XX in subject having genome with variant allele comprising cytochrome p450,
XX subfamily IIIA, polypeptide 5 polynucleotide, termed CYP3A5.
XX PS Disclosure; SEQ ID NO 683; 86pp; English.
XX CC The present invention describes the use of irinotecan (I) or its
XX derivative for the preparation of a pharmaceutical composition for
XX treating colorectal, cervical, gastric, lung, ovarian or pancreatic
XX cancer, or malignant glioma in a subject having a genome with a variant
XX allele which comprises a cytochrome p450, subfamily IIIA (nifedipine
XX oxidase), polypeptide 5 (CYP3A5) polynucleotide (II). (I) and (II) have
XX cytostatic activity. The therapeutic applications of (I) is improved,
XX since it is possible to individually treat a subject with an appropriate
XX dosage and/or an appropriate derivative of (I). Therefore, undesirable,
XX harmful or toxic effects are efficiently avoided. Unnecessary and
XX potentially harmful treatment of those subjects who do not respond to the
XX treatment with substances (nonresponders), as well as the development of
XX drug resistances due to suboptimal drug dosing can be avoided. ACF62200
XX to ACF62751 and ABW34912 to ABW35013 represent sequences used in the
XX exemplification of the present invention
XX SQ Sequence 177380 BP; 58338 A; 32530 C; 31389 G; 55123 T; 0 U; 0 Other;

Query Match 3.3%; Score 36.2; DB 7; Length 177380;
Best Local Similarity 50.9%; Pred. No. 17;
Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 182 TATATGGCTCTGGCGATGAGTAGCAATTATGAGCGATGCACGATGCTCTACTGTGC 241
DB 47048 TTTATGTGTAGCTCTCTGTGTGGACTTCAGCCGTCATGTACGATTTTCATCCACGATAC 46989

QY 242 TCTGTGTGTTAGGTTACCTTAGCTAGACAATATCACATACAAAATGTGTTTCCACATG 301
DB 46988 TTTAAGCCTTATGTTTATATATGCTGTGGAGGAGGAAAACAAAACGTTTGTCTCCATT 46929

QY 302 TCAGCTGTTCTACCGTAGTCTGAGTGAATGGGTAAFTTGAATATATGA 350
DB 46928 TCAGAGAGAGTAAATTAATCTGGGATAAAAAAGCCCTATTAAATTGATTCA 46880

RESULT 9
ADB20870/c
ID ADB20870 standard; DNA; 177380 BP.
XX AC ADB20870;
XX AC ADB20870;
XX 20-NOV-2003 (first entry)
DT MRPI based cancer related nucleic acid SEQ ID NO:683.
DE

```

KW irinotecan; colorectal cancer; cervical cancer; gastric cancer;
KW lung cancer; ovarian cancer; pancreatic cancer; malignant glioma;
KW variant allele; multidrug resistance protein 1; MRP1; Cytostatic; Gene;
XX ds.
XX
XX
XX Unidentified.
XX
XX WO2003013533-A2.
XX
XX 20-FEB-2003.
XX
XX 23-JUL-2002; 2002WO-EP008200.
XX
XX 23-JUL-2001; 2001EP-00117608.
XX
XX 24-MAY-2002; 2002EP-00011710.
XX
XX (EPID-) EPIDAUROS BIOTECHNOLOGIS AG.
XX
XX Heinrich G, Kerb R;
XX WPI; 2003-354397/33.
XX
XX Use of irinotecan or its derivative for preparation of a pharmaceutical
PT composition for treating cancer in a subject having a genome with a
PT variant allele comprising a multidrug resistance protein 1
PT polynucleotide.
XX
XX Disclosure; SEQ ID NO 683; 100pp; English.
XX
XX The present invention describes a method for the use of irinotecan (I) or
CC its derivative for the preparation of a pharmaceutical composition for
CC treating colorectal, cervical, gastric, lung, ovarian or pancreatic
CC cancer, or malignant glioma in a subject having a genome with a variant
CC allele which comprises a multidrug resistance protein 1 (MRP1)
CC polynucleotide (II). (I) has cytostatic activity. (I) or its derivative
CC can be used for the preparation of a pharmaceutical composition for
CC treating colorectal, cervical, gastric, lung, ovarian or pancreatic
CC cancer, or malignant glioma in a subject, where the subject is a human
CC (preferably African or Asian) or a mouse. The present sequence represents
CC a sequence which is used in the exemplification of the present invention.
XX
XX Sequence 177380 BP; 58338 A; 32530 C; 31389 G; 55123 T; 0 U; 0 Other;
SQ

Query Match 3.3%; Score 36.2; DB 7; Length 177380;
Best Local Similarity 50.9%; Pred. No. 17;
Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 182 TATATGGCTCTGGCGATGAGTAGCATTATAGGAGGCGATGCACACATGGCTCTACTGCTGC 241
Db 47048 TTTATGTGTAGCTCTCTGTGTGGACTTCAGCGTCATGTACGATTTTCATCCACAGTAC 46989
QY 242 TCTGTGGTTAGGTACCTTAGCTAGACAAATATCAATACAAATGGTGTTCACATG 301
Db 46988 TTTAAGCCTTATGTTAATAATGCTGTGGAAGGAAACAAACAGTTTGTCTCCATT 46929
QY 302 TCAGCTGGTTCTACCGTAGTCTGAGTGAATGGGTAATTTGATATATGA 350
Db 46928 TCAGAGAGAGTAAATTAATCTGGGATAAAAGCCCTATTAAATTTGATTC 46880

RESULT 10
ADB87959/c
ID ADB87959 standard; DNA; 177380 BP.
XX
XX ADB87959;
XX
XX 04-DEC-2003 (first entry)
XX
XX Human UGT1A1 gene sequence SEQ ID NO:683.
XX
XX irinotecan; cancer; UGT1A1; cytostatic; topoisomerase I inhibitor;
KW colorectal cancer; cervical cancer; gastric cancer; lung cancer;
KW ovarian cancer; pancreatic cancer; malignant glioma;
KW

KW uridine diphosphate glycosyltransferase1 member A1; gene; ds.
XX Homo sapiens.
XX
XX WO2003013536-A2.
XX
XX 20-FEB-2003.
XX
XX 23-JUL-2002; 2002WO-EP008217.
XX
XX 23-JUL-2001; 2001EP-00117608.
XX
XX 24-MAY-2002; 2002EP-00011710.
XX
XX (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX
XX Heinrich G, Kerb R;
XX WPI; 2003-289896/28.
XX
XX Use of irinotecan to treat cancer patient by determining if patient has
PT variant alleles of UGT1A1 gene, administering increased/decreased amounts
PT of irinotecan based on increased/decreased levels of UGT1A1 gene product.
XX
XX Disclosure; SEQ ID NO 683; 107pp; English.
XX
XX The invention relates to the novel use of irinotecan to treat a patient
CC suffering from cancer. This involves determining if the patient has one
CC or more variant alleles of the UGT1A1 gene, and if the patient has one or
CC more of such variant alleles, irinotecan is administered in an increased
CC or decreased amount in comparison to the amount that is administered
CC without regard to the patient's alleles in the UGT1A1 gene. The invention
CC has cytostatic activity. A composition of the invention acts as a
CC topoisomerase I inhibitor. The method is useful for treating a patient,
CC an animal e.g. mouse or a human, preferably African or Asian, suffering
CC from cancer such as colorectal, cervical, gastric cancer, lung, ovarian,
CC pancreatic cancer or malignant glioma. The present sequence is udes in
CC the exemplification of the invention.
XX
XX Sequence 177380 BP; 58338 A; 32530 C; 31389 G; 55123 T; 0 U; 0 Other;
SQ

Query Match 3.3%; Score 36.2; DB 9; Length 177380;
Best Local Similarity 50.9%; Pred. No. 17;
Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 182 TATATGGCTCTGGCGATGAGTAGCATTATAGGAGGCGATGCACACATGGCTCTACTGCTGC 241
Db 47048 TTTATGTGTAGCTCTCTGTGTGGACTTCAGCGTCATGTACGATTTTCATCCACAGTAC 46989
QY 242 TCTGTGGTTAGGTACCTTAGCTAGACAAATATCAATACAAATGGTGTTCACATG 301
Db 46988 TTTAAGCCTTATGTTAATAATGCTGTGGAAGGAAACAAACAGTTTGTCTCCATT 46929
QY 302 TCAGCTGGTTCTACCGTAGTCTGAGTGAATGGGTAATTTGATATATGA 350
Db 46928 TCAGAGAGAGTAAATTAATCTGGGATAAAAGCCCTATTAAATTTGATTC 46880

RESULT 11
ADB96942/c
ID ADB96942 standard; DNA; 177380 BP.
XX
XX ADB96942;
XX
XX 04-DEC-2003 (first entry)
XX
XX Human MDR1 related DNA sequence SEQ ID NO:683.
XX
XX irinotecan; colorectal cancer; cervical cancer; gastric cancer;
KW lung cancer; ovarian cancer; pancreatic cancer; malignant glioma;
KW multidrug resistance 1; MDR1; cytostatic; human; Cyp3A5; MRP1; MDR1;
KW TOP1; ds.
XX
XX Homo sapiens.
OS

DR WPI; 2002-154757/20.
 XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligonucleotides,
 PT useful for detecting cytosine methylation state of genes associated with
 PT pharmacogenomics and for therapy of diseases e.g. cancer.
 XX
 PS Claim 1; SEQ ID NO 141; 24pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence at least 18
 CC bases in length of a segment of the chemically pretreated DNA of genes
 CC associated with pharmacogenomics according to one of the sequences of the
 CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B (NM_000497), CYP3A3
 CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
 CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996),
 CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899, and
 CC their complementary sequences, or a sequence (S1) chosen from 87
 CC sequences and their complements. The chemical pretreatment is bisulphite
 CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
 CC Also included are an oligomer (Ii) in particular an oligonucleotide or a
 CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
 CC base sequence having a length of 9 nucleotides which hybridises to or is
 CC identical to a chemically pretreated DNA of genes associated with
 CC pharmacogenomics and their complements, arranged in an array for
 CC analysing diseases associated with the methylation state (CpG) and/or
 CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
 CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
 CC and their complements is useful for diagnosis and therapy of solid
 CC tumours and cancer. The present sequence represents one the 87 DNA
 CC sequences or its complement. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 7025 BP; 2377 A; 20 C; 1169 G; 3458 T; 0 U; 1 Other;
 Query Match 3.3%; Score 35.4; DB 6; Length 7025;
 Best Local Similarity 61.3%; Pred. No. 5.6;
 Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
 QY 536 TGTCATGCTATATGAATGAATGAAGAGAGGTGTGACGGGTAGTACGAATCTGTATGA 595
 DB 3990 TTTTACGTTATATTTTAAAGAAAATAAATAATGTTTATAGAAGTATGTATTAATGT 4049
 QY 596 TAATCAATGTCAGTCATGCTGTAACAGCGGA 628
 DB 4050 TAATTAACGTTATTTATATGTTTAAAGTGA 4082
 RESULT 15
 AAH16164/c
 ID AAH16164 standard; cDNA; 2604 BP.
 AC AAH16164;
 XX
 XX 26-JUN-2001 (first entry)
 DT
 XX Human cDNA sequence SEQ ID NO:14937.
 DE
 XX

PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 PA (EPIG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-010834/01.
 XX
 DR New nucleic acid, useful for diagnosis and therapy of metabolic disease,
 PT solid tumor and cancers, comprises segment of chemically modified genomic
 PT sequences of genes associated with metabolism.
 XX
 PS Claim 1; Page 119-121; 143pp; English.
 XX
 CC The invention relates to a nucleic acid (I) comprising a sequence at
 CC least 18 bases of a segment of the chemically pretreated DNA of genes
 CC associated with metabolism such as DUSP2 (NM_004418), EPHX2 (NM_001979),
 CC QDPR (NM_000320), SGSH (NM_000199), SHMT2 (NM_005412), SLC7A2
 CC (NM_003046), SLC7A4 (NM_004173) and TYMS (NM_001071) (all undefined). (I)
 CC are useful for diagnosis and therapy of metabolic disease, solid tumours
 CC and cancers; as primer oligonucleotides for the amplification of DNA
 CC sequences; for detecting the cytosine methylation state and/or single
 CC nucleotide polymorphisms (SNPs) in a chemically treated DNA of genes
 CC associated with metabolism. An array of (i) is useful for ascertaining
 CC genetic and/or epigenetic parameters for the diagnosis and/or therapy of
 CC existing diseases or the predisposition to specific diseases by analysing
 CC cytosine methylations. The method involves chemically treating genomic
 CC DNA sample by a solution of bisulphite, hydrogen sulphite or disulphite
 CC such that cytosine bases which are unmethylated at the 5th-position are
 CC converted to uracil or another base which is dissimilar to cytosine in
 CC terms of hybridisation behaviour and amplifying fragments of the
 CC chemically pretreated genomic DNA. The genomic DNA is from cells or
 CC cellular components which contain DNA, sources of DNA comprising, for
 CC e.g. cell lines, biopsies, blood, sputum, stool, urine, cerebral-spinal
 CC fluid, tissue embedded in paraffin such as tissue from eye, intestine,
 CC kidney, brain, heart, prostate, lung, breast or liver, histologic object
 CC slides and their combinations. Genetic parameters are mutations in
 CC particular insertions, deletions, point mutations, inversions and
 CC polymorphisms of genes associated with metabolism and sequences further
 CC required for their regulation. Epigenetic parameters are in particular
 CC cytosine methylations and further chemical modifications of DNA bases of
 CC genes associated with metabolism. Further epigenetic parameters include
 CC for e.g. the acetylation of histones which correlates with DNA
 CC methylation. AAS63306-AAS63373 represent chemically pretreated metabolism
 CC associated genes, and related primers of the invention
 XX
 SQ Sequence 7025 BP; 2377 A; 20 C; 1169 G; 3458 T; 0 U; 1 Other;
 Query Match 3.3%; Score 35.4; DB 6; Length 7025;
 Best Local Similarity 61.3%; Pred. No. 5.6;
 Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
 QY 536 TGTCATGCTATATGAATGAATGAAGAGAGGTGTGACGGGTAGTACGAATCTGTATGA 595
 DB 3990 TTTTACGTTATATTTTAAAGAAAATAAATAATGTTTATAGAAGTATGTATTAATGT 4049
 QY 596 TAATCAATGTCAGTCATGCTGTAACAGCGGA 628
 DB 4050 TAATTAACGTTATTTATATGTTTAAAGTGA 4082
 RESULT 14
 AAS63350
 ID AAS63350 standard; DNA; 7025 BP.
 XX
 AC AAS63350;
 XX
 XX 29-JAN-2002 (first entry)
 DT
 XX Chemically pretreated metabolism associated gene #45.
 DE
 XX Human; cytostatic; anti-tumour; metabolism; metabolic disease; liver;
 KW solid tumour; cancer; cytosine methylation; epigenetic; eye; kidney;
 KW single nucleotide polymorphism detection; SNP; stool; urine; lung;
 KW cerebral-spinal fluid; intestine; brain; heart; prostate; breast; DUSP2;
 KW EPHX2; QDPR; SGSH; SHMT2; SLC7A2; SLC7A4; TYMS; ds.
 XX
 OS Homo sapiens.
 XX
 XX WO200176451-A2.
 PN
 XX 18-OCT-2001.
 PD
 XX 06-APR-2001; 2001WO-EP004016.
 XX
 XX 06-APR-2000; 2000DE-01019058.

Job time : 337.836 secs

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-00116126.

29-JUL-1999; 99JP-00248036.

27-AUG-1999; 99JP-00300253.

11-JAN-2000; 2000JP-00118776.

02-MAY-2000; 2000JP-00182767.

09-JUN-2000; 2000JP-00241899.

(HELI-) HSEIX RES INST.

Ota T, Isogai T, Nishikawa T, Kayaishi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Claim 8; SEQ ID NO 14937; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 3'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2604 BP; 727 A; 491 C; 558 G; 828 T; 0 U; 0 Other;

Query Match 3.2%; Score 35.2; DB 4; Length 2604;

Best Local Similarity 53.7%; Pred No. 3.9;

Matches 73; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 514 ACAGTGTCTGCTGTGAACATGTCATGTCATATGAATGAATGAAGAGGTGTGAC 573

Db 1188 ACAGTGTCTGCTGTGAACATGTCATGTCATATGAATGAATGAAGAGGTGTGAC 1129

QY 574 GGGTTAGTACGAATCTGTATGATCAATGTCAGTGTGTCAGTGTGTCAGTGTGTCAGT 633

Db 1128 GGAATTAATGAAGAGGTGTGTCAGTGTGTCAGTGTGTCAGTGTGTCAGTGTGTCAGT 1069

QY 634 ATCTAGCAGTCTGATG 649

Db 1068 ATCAAGAACTCTGATG 1053

Search completed: June 10, 2004, 09:56:07

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 08:11:45 ; Search time 2997.29 Seconds
(without alignments)
15747.756 Million cell updates/sec

Title: US-10-070-386-2
Perfect score: 1089
Sequence: 1 taaactccatctatagcgg.....actgcaatgagtcgcccggg 1089

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_ma.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rnd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sv.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1089	100.0	1089	6	BD013074	Regulator
C 2	47	4.3	2000	6	AX655393	Sequence
C 3	44.8	4.1	2000	6	AX655393	Sequence
C 4	43	3.9	190760	2	AC109734	Rattus no
C 5	43	3.9	214458	2	AC117338	Rattus no
C 6	42.6	3.9	399	6	AR417479	Sequence
C 7	42.6	3.9	399	6	BD113032	EST and e
C 8	42.6	3.9	274651	2	AC095227	Rattus no
C 9	41.8	3.8	14461	6	AX256438	Sequence
C 10	41.8	3.8	128683	2	AC135793	Oryza sat
C 11	41.2	3.8	10732	6	E32986	Gene encodi
C 12	40.8	3.7	128683	2	AC135793	Oryza sat
C 13	40.4	3.7	220327	2	AC125966	Rattus no
C 14	40.4	3.7	24974	2	AC097706	Rattus no
C 15	40.2	3.7	1292	3	BT0307576	Eycturus t
C 16	40.2	3.7	75106	8	ATF16M2	Arabidops
C 17	40.2	3.7	113526	9	AL354769	Human DNA
C 18	40.2	3.7	174026	2	AC021864	Homo sapi
C 19	40	3.7	152304	2	BX072562	Danio rer
C 20	39.8	3.7	111477	9	AL451162	Human DNA
C 21	39.8	3.7	143678	9	AC005145	Homo sapi
C 22	39.8	3.7	170368	9	AC117518	Homo sapi
C 23	39.8	3.7	195156	9	AC090660	Homo sapi
C 24	39.8	3.7	198127	2	AF001593	Homo sapi
C 25	39.6	3.6	1141	6	AX083744	Sequence
C 26	39.4	3.6	175374	2	AC129987	Rattus no
C 27	39.4	3.6	254830	2	AC095718	Rattus no
C 28	39.2	3.6	91853	9	AC106810	Homo sapi
C 29	39.2	3.6	112445	9	AL590072	Human DNA
C 30	39	3.6	139189	10	AL935138	Mouse DNA
C 31	39	3.6	223545	2	AC100545	Mus muscu
C 32	39	3.6	318620	2	AC110910	Mus muscu
C 33	38.8	3.6	37797	3	CFT21B4	Caenorhabdi
C 34	38.8	3.6	125020	9	AF429315	Homo sapi
C 35	38.6	3.5	16130	2	AC013873	Drosophil
C 36	38.6	3.5	41609	3	AF018451	Caenorhab
C 37	38.6	3.5	164761	3	AC023695	Drosophil
C 38	38.6	3.5	198245	3	AC023701	Drosophil
C 39	38.6	3.5	198966	2	AC120873	Mus muscu
C 40	38.6	3.5	319329	3	AE003432	Drosophil
C 41	38.4	3.5	162097	2	EX284614	Danio rer
C 42	38.4	3.5	167531	5	AL529135	Zebrafish
C 43	38.4	3.5	185624	2	EX322551	Danio rer
C 44	38.4	3.5	214843	5	EX005203	Zebrafish
C 45	38.2	3.5	225020	10	AC109247	Mus muscu

ALIGNMENTS

RESULT 1
BD013074
LOCUS
DEFINITION BD013074 1089 bp DNA linear PAT 02-AUG-2002
Regulatory sequences and expression system functional in mold
fungi
ACCESSION BD013074
VERSION BD013074.1 GI:22093263
KEYWORDS WO 0118219-A/2.
SOURCE unidentified
ORGANISM unclassified
REFERENCE 1 (bases 1 to 1089)
AUTHORS Watanabe,M. and Murakami,T.
TITLE Regulatory sequences and expression system functional in mold fungi
JOURNAL Patent: WO 0118219-A 2 15-MAR-2001;

Pred. No. is the number of results predicted by chance to have a

Db	789	WRCWAGNARWMSRVRWKKYATRYVWVWNTWWSVRWKSVRWWSGMRGZWSAW	730		
Qy	542	GTCTATATGAATGAAGAAGAGGTGTGACGGGTAGTACGAATCTGTATGATATCA	601		
Db	729	RYCSRKKCAKTKYASSARVTKARRSYRVRWYMRKGMWYRYSRMTFRAMSR	670		
Qy	602	ATGGTAGCAGTGTGTAACA---GCGGATCGGATCTAGCACTGCTATGCTGGTA	657		
Db	669	RWAGASMSKWNWYRGASWYKYSKACCKRYMTSSYKSMYGMVSSYSNSW	610		
Qy	658	TGTAATCTCGGTATGTTCAAGGCGCATAGAAGAACTCCTCAGTGTGAGCATCG	717		
Db	609	TSKSYMGKMTCTMTYSKMSGSTRSPKGRWSGMSRMYMRWKKRKYRMYMKWCTWR	550		
Qy	718	TAAGTCTGTACATTTCACTGCAATTTCTGAACAATTCGAGAGCATTTAGAACTAA	777		
Db	549	RCMYRWGTYMTTISRMYTORYKARTSKRYMYKTRKYCTWYTYTGMYTKCSIMR	490		
Qy	778	ATGAACTCTCTATTAAGTGAAGAACAGAGCGCCCTTTTATTAAGAAACAGAGCGTC	837		
Db	489	YGYCKACKCCYACWCAAYSGMMWYRKYSKWMBSTKXMMWYKCKRSMKYGAKC	430		
Qy	838	AAGAAGCTCTTCAAGCTCATCAGAGCGGTTCCATCCAGATCATCTTCCCTTGAACA	897		
Db	429	YCKMWTYCSYGMWYTYGSKYKRCYKYGMYKGMWYMYKGMWYMYSSMMTYTYTA	370		
Qy	898	TGTTCTCGCATTCAGATCGTACGATGAAACCGTCCAGGCTGCTCTCATCTCCCTG	957		
Db	369	KYWKYFRGMSWYKSKYKCYCTWYKCMRKYRWRKWKYKSKYKCYWRYAT	310		
Qy	958	CGTCCCTCGCAATAAATCGTATTACCA-TTTTCTTCCGAGCGCGGTCAAGTGAGCG	1016		
Db	309	CYWCCTYRGRYSRSMRTAGMKWRWSRWSRWSYKMYKGMWYKSKYKMYKGMWYK	250		
Qy	1017	ACGTGCCAGCTGGAGTCACAAATCAGCATGATCGTCATCCAGGCC	1065		
Db	249	SGTWSRAAKRYKGYSTSRRAKWRACRMYKACRYKSTSYCCGSCYC	201		
RESULT 3	AX655393	2000 bp	DNA	linear	PAT 22-MAR-2003
LOCUS	Sequence 5263 from Patent WO03000898.				
DEFINITION	AX655393				
ACCESSION	AX655393.1	GI:29158207			
VERSION					
KEYWORDS	Oryza sativa				
SOURCE	Oryza sativa				
ORGANISM	Oryza sativa				
REFERENCE	Chang, H. S., Chen, W., Cooper, B., Glazebrook, J., Goff, S. A., Hou, Y. M., Katagiri, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G. G.				
AUTHORS	Plant genes involved in defense against pathogens				
TITLE	Patent: WO 03000898-A 5263 03-JAN-2003;				
JOURNAL	Syngenta Participations AG (CH)				
FEATURES	Location/Qualifiers				
source	1. 2000				
ORIGIN	/organism="Oryza sativa"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:4530"				
Query Match	4.1%; Score 44.8; DB 6; Length 2000;				
Best Local Similarity	11.3%; Pred. No. 0.06;				
Matches 104; Conservative 398; Mismatches 410; Indels 12; Gaps 4;					
Qy	18	CGGTGGCTGAAAGGAGGAGGAGGAGATCAGTCTTTTGACCACTGGAGATCT	77		
Db	73	SGRTSGSSGSGYSGKXKRYKRSKRWGRGRGRGRGRGRGRGRGRGRGRGRGRGR	132		

Qy	78	GATCTGTGTTTCAGTCTCAGTAACATCTGTGTGGAAGTTACATCTTCTGGTCTCCTCTTAC	137
Db	133	GRMMGGKSRMSYMMCYARGCGSKRKKSGSGWKTCKRGARGGSGSSGAKYKSGSHS	192
Qy	138	CAGCCCTCCAGCCACCAACAGATGTTAGAGTTTCGCTCATTTATATGCTCTGGCGA	197
Db	193	KBMWSSCGRSRGSRSAYSRVYGTSRKYVYKMTYYSASRCMAYMTTSYACSSYT	252
Qy	198	TGAGTAGCATTTATGAGGATGACAGCATGCGCTCTACTGCTCTCTGTGTGGTAGGTTA	257
Db	253	WCRSKRSRMMKMRKMRWSRSYGSYKMMCTATYKKSYSRWKCYWYRGGWGRATR	312
Qy	258	CCTTAGCTAGACAATATCAACAATAAATGTGTTTCCACATGTGACGTGGTCTTACCG	317
Db	313	YMGRCYMSBAMMYKMYVYRGYKMGKRWAGMMMSKCRWSKACYWYRMRWRMTTR	372
Qy	318	TAGTCTGAGTGAATGGTAAATGATATATTGAGCTTGAACCCGCAATATTGTAACAGAG	377
Db	373	RWAKKSRSTSRKCRKWKMRKRYKMGYSGMRSCKRARWKRCKRSGRWMKMGRCGM	432
Qy	378	CCAAACAATGGGTCACTCGGCCCCACACATGTGGCTATATAAGCTACTGTCTA----	432
Db	433	TCBMSYGMWRKSKWBAWKYKWSRMYRWRKKSRTTMMKTRGGMMGTGRCRYK	492
Qy	433	--GCAATCAGACTTACTGATAGAGCTCCCTATATGTCATAAAATAGTCACTACTAG	490
Db	493	RSGMKRCKRKRWRGRWRYRWRKRYTMYRKYKYSYSAARKARCMYRGKYWA	552
Qy	491	AACTACCGACAGTGTGAAATCGACAGTGTCTGGTCTGTTGAACATGTCTATGTCTATG	550
Db	553	GHMMKRYKMYWYKMMWYKRYKSKSWYKMSYASCMKSRKAGAKMKRCKMSAMSK	612
Qy	551	AATGAATAGAGAGGTGTGACGGGTAGTACGAATCTGTATGATATAATCAATGGTAGCA	610
Db	613	SMRSRCKRCKASKRSKRYAMMGMTSGSRMSRWKSY-TCYWRKWSMKSTCTMYM	671
Qy	611	GTGATGTGTAACAGCGGATCGGATCTAGCACTGTCTATGCTGGGTATGTAATCTCGGT	670
Db	672	SKYTYAKYGSYWRVRYPAWCMYMRWYTYRYSYMTYMAWYSSRMAWTKMYSGRYWT	731
Qy	671	ATGTTCATAGGGGACATAGAAGAATACCTCAGTGTGACATAGTACGTCTGTA-C	729
Db	732	SWYKCKSKYKSMYTYWSWAKTKMWRRYATRMMWNYRYSKMYTWCTTMNGYWY	791
Qy	730	ATTTCACTGCATAATTTCTGAACAATTTGGAGAGCATTATGAATACTAATAATGGAATCTCT	789
Db	792	WRTYMKRYYMYKCTKYTYWSATYTYGTWAAWMAKTRWGMGTGAKTRGRARAYW	851
Qy	790	ATTATAGTGAAGAACAGAGCGCCCTTTTATTATGAAGACAGAGCGCTCAAGAACGCTTT	849
Db	852	WKMATWCAATKRWMTKKGAK---WAWTWAKAWRYKYSWMAWYTYKTRTRYKTCW	908
Qy	850	CAAGCTCATCAGAGCGGTTCATCCAGATCATACTTCCCTTGAACCATGTTCTCCCAT	909
Db	909	KARWGSWAYWMMWKSASAKMMWMMKGGRWGTWYWYCTTWKACGRATKYMCCAGWA	968
Qy	910	CAGAATCGTAGCATGGAACCGT	933
Db	969	MYSYSWTRTYMRTWRMMWASSRT	992
RESULT 4			
AC109734/c			
LOCUS	Rattus norvegicus clone CH230-324E22, ***	190760 bp	DNA linear HTG 09-OCT-2002
DEFINITION	***, 2 unordered pieces.		
ACCESSION	AC109734		
VERSION	AC109734.4 GI:23195415		
KEYWORDS	HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

REFERENCE
AUTHORS

1 (bases 1 to 190760)
Muzny, D., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, P., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Danson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, I., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Gracy, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Hawlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, D., Kovar, C., Lewis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulsegod, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, B., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Pu, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, W., Richards, S., Riggs, P., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Steed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valdes, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.

Direct Submission
Unpublished
2 (bases 1 to 190760)
Worley, K. C.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL

Direct Submission
Submitted (07-PEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 190760)
Rat Genome Sequencing Consortium.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL

Direct Submission
Submitted (09-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 19, 2002 this sequence version replaced gi:21738234.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: QGMU
Center clone name: CH230-324E22
Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 169268 bases at least Q40
Consensus quality: 172261 bases at least Q30
Consensus quality: 173949 bases at least Q20
Estimated insert size: 209466; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: this is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 180102: contig of 180102 bp in length
1 180103 180202: gap of unknown length
1 180203 190760: contig of 10558 bp in length.

FEATURES
source
1..190760
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-324E22"
1..1725
/note="wgs end extension
clone_end:Sp6"
2044..2981
/note="clone boundary
clone_end:Sp6"
site:MboI
179212..180102
/note="clone boundary
clone_end:T7"
site:MboI
end_sequence:EXANV35TJ"
180203..181221
/note="wgs end extension
clone_end:T7"

misc_feature
652 TGGGATGTAATCTCGGTATGTTCTATGAAGGGGCGACATAGAAAGAAATACCTCAGTGTGAG 711
71698 TGTGTATTTTCATAGCTCTCTTGTGAAATATTTTGTTCAGAAACAGAGAACTAAATATAAG 71639

misc_feature
712 CATACGTAAGCTCTGTACATTTCTCAATTTCTGAAATTTCTGAACTTCGACAGCATTTATGAAA 771
71638 AGTTCCCTTATTAACCTGGGCCATCTCTGAAGATTTGATTAATGATAGACACAGAT 71579
772 TACTAAATGGAACCTCCTCATTATAGTGGAAAACAGAGCGCCCTTTTATTATGAAACAGA 831
71578 TATTAGCATGATGATCAACATGATATAAGCAAAAACCAACCAAGCTTTCAGTTACAAACCA 71519

Query Match 3.9%; Score 43; DB 2; Length 190760;
Best Local Similarity 48.9%; Pred. No. 0.43;
Matches 115; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 652 TGGGATGTAATCTCGGTATGTTCTATGAAGGGGCGACATAGAAAGAAATACCTCAGTGTGAG 711
Db 71698 TGTGTATTTTCATAGCTCTCTTGTGAAATATTTTGTTCAGAAACAGAGAACTAAATATAAG 71639

QY 712 CATACGTAAGCTCTGTACATTTCTCAATTTCTGAAATTTCTGAACTTCGACAGCATTTATGAAA 771
Db 71638 AGTTCCCTTATTAACCTGGGCCATCTCTGAAGATTTGATTAATGATAGACACAGAT 71579
QY 772 TACTAAATGGAACCTCCTCATTATAGTGGAAAACAGAGCGCCCTTTTATTATGAAACAGA 831
Db 71578 TATTAGCATGATGATCAACATGATATAAGCAAAAACCAACCAAGCTTTCAGTTACAAACCA 71519

Direct Submission
Submitted (10-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 214458)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Oct 9, 2002 this sequence version replaced gi:21746250.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GUAX
Center clone name: CH230-351P22
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 145461 bases at least Q40
Consensus quality: 146349 bases at least Q30
Consensus quality: 147003 bases at least Q20
Estimated insert size: 156806; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: this is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 212885: contig of 212885 bp in length
* 212886 212985: gap of unknown length
* 212986 214458: contig of 1473 bp in length.

Location/Qualifiers
1. 214458
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-351P22"
1. 1014
/note="wgs end extension
clone end:Sp6"
39152_40502
/note="wgs end extension
clone end:Sp6"
44260_45151
/note="clone boundary
clone end:Sp6
site:Mbol
end sequence:RXASL95TV"
complement(210250..210949)
/note="clone boundary
clone end:T7
site:Mbol

832 AGCGTCAAGACGCTTTCAACGTCATCAGAGCGGTCATCCAGATCATACTTT 886
Db 71518 TGTTCCTGTAAGTGAAGTCTGAGGCGATTCATCTCTTTGCTTCAGACCAACTAT 71464

AC117338 214458 bp DNA linear HTG 11-OCT-2002
Rattus norvegicus clone CH230-351P22, *** SEQUENCING IN PROGRESS
***, 2 unordered pieces.
AC117338 3 GI:23617965
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
1 (bases 1 to 214458)
Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, I., Garza, M., Geunargenis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hu, Y. K., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, D., Kovari, C., Kowis, C., Kratt, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulsegod, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, B., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, B., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munday, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoelameh, O., Okwuonu, G., Olarnpungagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, P., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Fuazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Shetty, A., Sodargren, E., Song, X. Z., Sorelle, R., Soosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, L., Trejos, Z., Umani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G., and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 214458)
Worley, K. C.

```

end_sequence:RXASL95TJ"

Query Match      3.9%; Score 43; DB 2; Length 214458;
Best Local Similarity 48.9%; Pred. No. 0.43;
Matches 115; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 652 TGGGTATGTAATCCTCGGTATGTTTCATAGGGCGACATAGAAAGAAATACCTCAGTGTACG 711
Db 63587 TGTGTATTTTCATAGCTCTCTTGTAATATATTTGGTCAGAAACAGAGAACTAAATATAAG 63646

QY 712 CATACGTAGCTCTGTACATTTTCATCTGCAAAATTTCTGAAACAATTTGGAGAGCATTTACAA 771
Db 63647 AGTTCCTTTAAATCGGCCATCTCATGAAGATTGATTGATAAATGATAAGACACAAGAAAT 63706

QY 772 TACTAAATCGAACTCTCATTTATAAGTGGAAACACAGAGCGCCCTTTTATTATGAAACAGA 831
Db 63707 TATTAGGCATGAGTCAACATGATTAAGACAAAACCAACGCTTTCAGTTGACGTTACAAACCA 63766

QY 832 AGCGTCAAGAAAGCTCTTTCAACGTCATCAGAGGCGTTCCATCCAGATCATACTTT 886
Db 63767 TGTTCCTTGAAGTCACTCAGCGGATTCATCTCTTTTCCCTCAGACCAACTAT 63821

RESULT 6
LOCUS AR417479 399 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 8976 from patent US 6639063.
ACCESSION AR417479
VERSION AR417479.1 GI:40172589
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 399)
AUTHORS Edwards,J.-B.D.M., Jobert,S. and Giordano,J.-Y.
TITLE EST's and encoded human proteins
JOURNAL Patent: US 6639063-A 8976 28-OCT-2003;
FEATURES
location/Qualifiers
1..399
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match      3.9%; Score 42.6; DB 6; Length 399;
Best Local Similarity 9.6%; Pred. No. 0.22;
Matches 33; Conservative 164; Mismatches 148; Indels 0; Gaps 0;

QY 740 AATTTCTGAACAATTCGAGAGCATTAATAACTAAATGGAACCTCCTCATTTATAGTG 799
Db 16 RANTRAYWGYTRGSKKSWRAKSMWKRRRRRAWWKSMCKWKKSKSWKSWGMMTK 75

QY 800 GAAACACAGAGCCCTTTTATTATGAACACAGAGCGTCAAGAAAGTCTTTCAACGTCATC 859
Db 76 RMKGRGAASWAGYMSWMTYTRWRYRYRYKCACTKWRAAAGMNGKAGWAWAYAKWYMA 135

QY 860 AGAGCGGTTCCATCCAGATCATACTTTCCCTTGAACCATGTTCTCGCATTCAGAACTGTA 919
Db 136 WRTANKYWMWKKKSRERRRRRAWWYMYMARTTWGHRASCYRAYVASAGYMYWM 195

QY 920 GCGATGGAACCGTCCAGGGTTCCTGTATTCCTCGGTCCCTTCGCAATAAAATCGTA 979
Db 196 YTMERKMWYSAGWSMKEKWRRCASYSWCSYCWGAKMMYKTSRWSYSSYRCTX 255

QY 980 TTACCATTTTCTTCGACGCGCGGTCAACTCGTACGACGCGTGCCACGTTGGAGTCCACA 1039
Db 256 YRSCCWSMSCKYNTYRYSWVASCYSYKTKBASCOMCCCKWRKMMWMMYMKTY 315

QY 1040 ATGACCATGGATTCGTATCCACGCCACTCACTCGCAATGAGTCG 1084
Db 316 CKKTSAMRYRSCYSAKRRWRWMMCCAYMMKTCSSWMMWMCWTGGC 360

```

AC095227.4 GI:22773024
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 274651)
Kuzny,D.Marie, Metzker,M.Lee., Abranzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angulano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenship,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,B.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chaves,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,I., Foster,M., Foster,P.,
Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgievski,B., Geis,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Loretsuwha,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwakoleme,O., Okwono,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfaunkoch,C.,
Plopper,P., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puaoro,M., Quiroz,J., Rachin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajds,D.,
Sneed,A., Sodargren,B., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villanana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleciyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,P., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G., and Gibbs,K.A.
Direct Submission
Unpublished
2 (bases 1 to 274651)
Worley,K.C.
Direct Submission
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 274651)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 10, 2002 this sequence version replaced gi:21722848.
The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GDKW
Center clone name: CH230-9G5
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 239968 bases at least Q40
Consensus quality: 232929 bases at least Q30
Consensus quality: 234953 bases at least Q20
Estimated insert size: 261923; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* {see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html}
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 136946: contig of 136946 bp in length
* 136947 137046: gap of unknown length
* 137047 143298: contig of 12252 bp in length
* 143299 143398: gap of unknown length
* 143399 168602: contig of 19204 bp in length
* 168603 168702: gap of unknown length
* 168703 265979: contig of 97277 bp in length
* 265980 267699: gap of unknown length
* 267699 267730: contig of 1290 bp in length
* 267730 270513: gap of unknown length
* 270513 270614: contig of 3044 bp in length
* 270614 274651: contig of 4038 bp in length.
* 274651 Location/Qualifiers
* source
* 1. 274651
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-9G5"
* 1. 1640
* /note="wgs end extension
* clone end:Sp6"
* 9989_10687
* /note="clone boundary
* clone end:Sp6"
* site:EcoRI
* end_sequence:BH341131"
* 10005_10709
* /note="clone boundary
* clone end:Sp6"
* site:EcoRI
* end_sequence:BH341131"
* misc_feature
* 3.94; Score 42.6; DB 2; Length 274651;
* Best Local Similarity 51.34; Pred. No. 0.6;
* Matches 99; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
* Query Match
* 640 CACTGCTATGCTGGGTATGTAATCTCGGTATGTTCTATAGGGCGCATAGAAAGATA 699

TITLE Direct Submission
JOURNAL Submitted (04-NOV-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
COMMENT On Nov 4, 2003 this sequence version replaced gi:24431631.
* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and the accession number will be preserved.
* The accession number will be 128683 bp in length.
FEATURES
Source Location/Qualifiers
1..128683
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="11"
/clone="OSJNB0017P15"
/note="japonica cultivar-group"
ORIGIN
Query Match 3.7%; Score 40.8; DB 2; Length 128683;
Best Local Similarity 11.9%; Pred. No. 1.9;
Matches 58; Conservative 219; Mismatches 207; Indels 4; Gaps 2;
QY 397 CCCCCAGACATGCTATATAGCTACTCTCTAGCAATCAGACTTACTGATGAGCG 456
DB 50701 YCYKSMKMKMSYAYRYMRSWYTSYMKAMYRSGWYWRKMAAYWYTGMYK 50642
QY 457 TCCCCCTATATGCTATAAATAGTCACTACTAGCACTACCGACAGTGTGAATCCGACA 516
DB 50641 WRKNGRGAMMYWCRWMAWYRCMTKRGTCGASWKMMSAASAAKYAWYCRMKRKKW 50582
QY 517 GTGCTGCTGCTGTGACATGCTATGCTATATGAT---GAATAGAGAGAGGTGAC 573
DB 50581 KASYTRMKWKRMAWMTYRHGRARWYTRHGRARWYTRHGRARWYTRHGRARWY 50522
QY 574 GGGTTAGTACGATCTGTATGATATCAATGCTAGCAGTGTGTAACAGCGGATCGG 633
DB 50521 KRRRTYMSASRKKMMWMSWRSRYWMSWMSWRSRYWMSWMSWRSRYWMSWMSW 50462
QY 634 ATCTAGCAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 693
DB 50461 MASATGKMWAGTKAKRSMTYRGAARWYTWYKWKWY-WGGTATKRYTWKWKRAW 50403
QY 694 AGAATACCTCAGTCAGCAGCAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 753
DB 50402 TKTATYCYATATWCTTAGSMWNTYTKKCRWTRWAKWSYTABSYTKRSKX 50343
QY 754 TTGGAGAGCATTAATGAATGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 813
DB 50342 KKRYKRYRMRKRYWARKWSYRRKRWYTWYKWKWYTWYKWKWYTWYKWKWYTWY 50283
QY 814 CTTTATTATGAACAGAGCGCTCAGACGCTTTCACGCTCAGAGCGGCTCCATC 873
DB 50282 YKMAWCCACATAAKASAYARRWTTWKKWMAATCCWYTWYKWKWYTWYKWKWY 50223
QY 874 CAGATCAT 881
DB 50222 CTAACCAT 50215
RESULT 13
AC125966/c 220327 bp DNA linear HTG 10-MAY-2003
LOCUS Rattus norvegicus clone CH230-2303, *** SEQUENCING IN PROGRESS ***,
DEFINITION 3 unordered pieces.
ACCESSION AC125966
VERSION AC125966.6 GI:30520551

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
AUTHORS 1 (bases 1 to 220327)
Murny,D,Waris,, Metzker,M, Lee,, Abramson,S,, Adams,C,, Alder,J,, Allen,C., Allen,H., Alsbrooks,S., Amin,A., Ayodeji,M., Baca,E., Baden,H., Baldino,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cadenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.D., Davis,C., Davy-Carroll,S., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,K., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escott,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Healand,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhwari,M., Louissegh,H., Lozano,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A., Mangun,B., Mapus,P., Martin,K., McNeill,T.Z., Meenen,E., McWhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K., Nwokileme,O., Okwona,G., Olapkinsagcon,A., Pal,S., Pankaj,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindestex,A., Popovic,D., Primus,E., Pu,L., L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczek,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 220327)
Worley,K.C.
Direct Submission
Submitted (02-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 220327)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:24942803.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 220327)
Murny,D,Waris,, Metzker,M, Lee,, Abramson,S,, Adams,C,, Alder,J,, Allen,C., Allen,H., Alsbrooks,S., Amin,A., Ayodeji,M., Baca,E., Baden,H., Baldino,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cadenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.D., Davis,C., Davy-Carroll,S., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,K., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escott,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Healand,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhwari,M., Louissegh,H., Lozano,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A., Mangun,B., Mapus,P., Martin,K., McNeill,T.Z., Meenen,E., McWhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K., Nwokileme,O., Okwona,G., Olapkinsagcon,A., Pal,S., Pankaj,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindestex,A., Popovic,D., Primus,E., Pu,L., L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczek,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 220327)
Worley,K.C.
Direct Submission
Submitted (02-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 220327)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:24942803.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas

(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRGV
Center clone name: CH230-2303
----- Summary Statistics

Assembly program: Atlas 3.0;
Consensus quality: 203711 bases at least Q40
Consensus quality: 206214 bases at least Q30
Consensus quality: 207905 bases at least Q20

Estimated insert size: 211787; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 216855: contig of 216855 bp in length
* 216956: gap of unknown length
* 216956 218421: contig of 1466 bp in length
* 218422 218521: gap of unknown length
* 218522 220327: contig of 1806 bp in length.

FEATURES

source
1. 220327
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-2303"

misc_feature

1. 1645
/notes="wgs end extension
clone_end:T7"

misc_feature

4726..6822
/notes="wgs end extension
clone_end:T7"

misc_feature

complement(7464..8057)
/notes="clone boundary
clone_end:T7"

misc_feature

site:ECORI
end_sequence:BH275795"
complement(215393..215875)

misc_feature

/notes="clone boundary
clone_end:Sp6
site:ECORI
end_sequence:BH275796"

ORIGIN

Query Match 3.7%; Score 40.4; DB 2; Length 220327;
Best Local Similarity 53.1%; Pred. No. 2.7;
Matches 86; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 211 TGAGGATGACACAGATGGCTACTGCTCTGTGTGGTTAGGTACCTTACCTAGACAGA 270
DB 14182 TCAGGCTCTGATACATAGTGTGGTGTAGCTGTACCCCTCATATCCATCCATACC 14123

QY 271 ATATCACAATACAAATGTTGTTCCACATGTCAGCTGGTCTTACCTAGTCTGAGTGAA 330
DB 14122 TAGGCAAAATAGTAATGTTGTTGACATGGCATGTTGATTAACCAACACAGAGCTTA 14063
QY 331 ATGGGTAATGATATATGAGCTTGACCCCGCAATATGTAA 372
DB 14062 TTGTCTCTGATGTTAGGCGAGCAATCTGCCCTATGATA 14021

RESULT 14

AC097706

LOCUS

DEFINITION

AC097706

AC097706.7 GI:305890908

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 244974)

REFERENCE

AUTHORS

AC097706 244974 bp DNA linear HTG 13-MAY-2003
Rattus norvegicus clone CH230-72A1, WORKING DRAFT SEQUENCE, 4
unordered pieces.

AC097706

AC097706.7 GI:305890908

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 244974)

REFERENCE

AUTHORS

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,

Allen, C., Ailen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyaiebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyie, M., Cree, A., D'Souza, L.,

DeValle, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Ducan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Gebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,

Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorensuhera, L., Louisedge, H., Lozano, R. J., Lu, X., Ma, J.,

Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,

Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E. E.,

Milosavljevic, A., Miner, G., Mirja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

Nwaelele, O., Okwuonu, G., Olarnpungoon, A., Pal, S., Parks, K.,

Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,

Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L.,

Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P.,

Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,

Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,

Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,

Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J.,

Steimle, M., Strong, N., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,

Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,

Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,

Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,

Williams, G., Willson, R., Wlecsyk, R., Wooden, H., Worley, K.,

Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V.,

Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von

Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,

Weinstock, G. and Gibbs, R. A.

Direct Submission

TITLE

```

Unpublished
2 (bases 1 to 244974)
Worley, K.C.
Direct Submission
Submitted (21-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 244974)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:22855886.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GIWD
Center clone name: CH230-72A1
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 229343 bases at least Q40
Consensus quality: 23227 bases at least Q30
Consensus quality: 234382 bases at least Q20
Estimated insert size: 240376; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 239157: contig of 239157 bp in length
* 239158 239257: gap of unknown length
* 239258 240297: contig of 1040 bp in length
* 240298 240397: gap of unknown length
* 240398 241731: contig of 1334 bp in length
* 241732 241831: gap of unknown length
* 241832 244974: contig of 3143 bp in length.
FEATURES
source
1. 244974
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-72A1"
1. .2094
misc_feature
/note="wgs contig"
135018..138051
/note="wgs_contig"
ORIGIN
Query Match 3.7%; Score 40.4; DB 2; Length 244974;
Best Local Similarity 53.1%; Pred. No. 2.8;

```

Search completed: June 10, 2004, 11:49:43
Job time : 3003.79 secs
